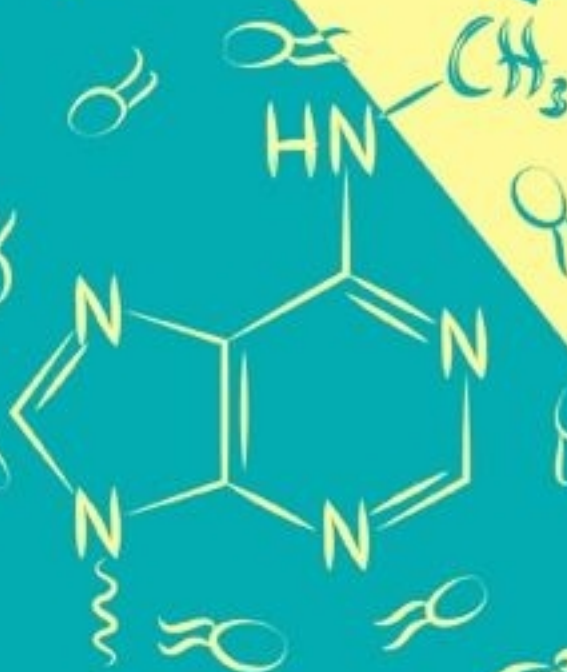


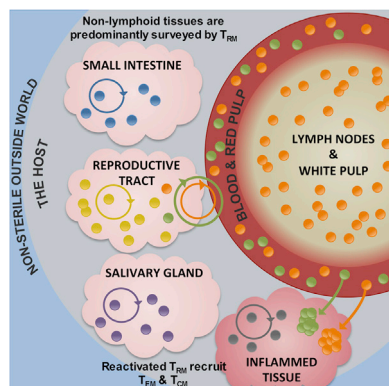
Cell

Volume 161
Number 4

May 7, 2015

www.cell.com





T Cells Join Neighborhood Watch

PAGE 737

Steinert et al. find that many anatomic compartments, including solid organs and vascular spaces, are patrolled by local resident populations of memory CD8⁺ T cells, concluding that immunosurveillance is far more regionalized than previously anticipated.

Metabolic Priming to a T

PAGE 750

Memory T cells protect against intracellular pathogens by scanning host cell surfaces and are critical for long-term immunity. Cui et al. describe how interleukin 7 rewires the metabolism of memory CD8⁺ T cells to allow them to import glycerol for triglyceride synthesis and storage to be able to sustain ATP levels for long-term metabolic fitness and rapid response to re-infection.

Non-Coding Code for Immunoglobulins

PAGE 762 and 774

Class-switch recombination (CSR) is essential for the generation of appropriate antibody responses to diverse stimuli. Zheng et al. discover that transcription through the switch region generates a non-coding RNA that directly binds to AID and guides the enzyme to DNA in a sequence-specific manner, positioning it in the right spots in the genome. In a separate study, Pefanis et al. identify a class of eRNAs that are targets of the RNA exosome machinery. When the RNAase activity of the exosome complex is ablated, eRNA-expressing regions accumulate deleterious R-loops. Notably, one of these eRNA-expressing elements, termed lncRNA-CSR, is required for long-range DNA interactions that regulate the immunoglobulin region super-enhancer function.

New tRFs in Cancer

PAGE 790

Upon stress, tRNAs are enzymatically cleaved, yielding distinct classes of tRNA-derived fragments. Goodarzi et al. find that some tRNA-derived fragments act as tumor suppressors through a post-transcriptional mechanism that leads to destabilization of many pro-oncogenic transcripts. Highly metastatic cells are capable of evading this mechanism by blunting the induction of tRFs during hypoxic conditions associated with cancer progression.

Sugar Rush for Good Vision

PAGE 817

The rod-derived cone viability factor RdCVF promotes survival of retinal cones, protecting them from neurodegeneration. Ait-Ali et al. now identify its cell-surface receptor and demonstrate that RdCVF binding accelerates glucose transporter function, enhancing the entry of glucose into photoreceptors and its oxidation through aerobic glycolysis. Exploiting a pathway that is also used by fast-dividing cancer cells, RdCVF maintains cone survival by mediating cone outer segment renewal.

A Structured Approach to Anti-Hypertensives

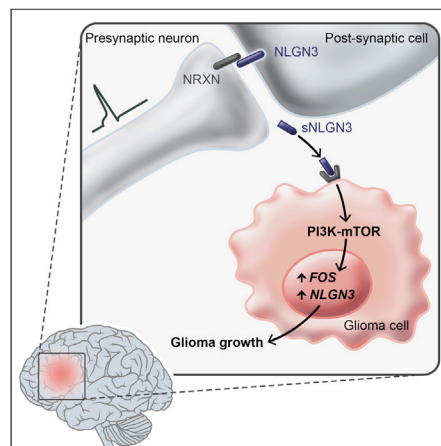
PAGE 833

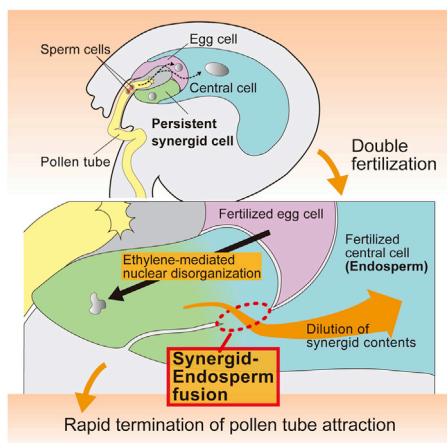
Angiotensin II type 1 receptor (AT₁R) is a G protein-coupled receptor regulating blood pressure. Using the recently developed method of serial femtosecond crystallography at an X-ray free-electron laser, Zhang et al. determine the room-temperature crystal structure of human AT₁R bound to an antagonist. Further docking analyses reveal the binding modes for other common anti-hypertensive drugs and in combination provide new insights into AT₁R structure-function relationships and for structure-based drug design.

Neural Boost for Gliomas

PAGE 803

Malignant gliomas are the leading cause of brain tumor deaths. Venkatesh et al. now find that neuronal activity promotes the growth of a broad range of pediatric and adult malignant gliomas through mechanisms that include activity-regulated secretion of the synaptic protein neuroligin-3. Secreted neuroligin-3 functions as a mitogen, recruiting the PI3K-mTOR pathway to induce glioma cell proliferation. Soluble neuroligin-3 also induces feed-forward glioma cell neuroligin-3 expression, which correlates inversely with survival in patients.





Sperm Counting 1, 2...

PAGE 907

In flowering plants, one sperm fertilizes an egg and another sperm fertilizes the central cell to form the embryo and endosperm, respectively. The synergid cells controlling the attraction of pollen tubes need to be inactivated after double fertilizations to block arrival of excess pollen tubes (polytubey). Maruyama et al. show that unusual cell fusion between the persistent synergid cell with endosperm and embryo-mediated elevation of ethylene signaling coordinately prevents polytubey.

Human Ribosome in Action

PAGE 845

Using multi-particle cryo-EM analysis, Behrmann et al. reveal 11 distinct functional states from a native actively translating human polysomal sample, providing insights into the configuration of the human ribosome at near-atomic resolution and highlighting the functional importance of both rigid and flexible regions.

Ancient Chaperone's Heme-ngous Role

PAGE 858

The mitochondrion maintains and regulates its proteome with chaperones primarily inherited from its α -proteobacterial ancestor. Kardon et al. find that one such mitochondrial chaperone, ClpX, directly stimulates a key heme biosynthetic enzyme by accelerating the incorporation of its cofactor, thereby controlling eukaryotic heme levels from yeast to vertebrate erythrocytes.

Ad-ding to DNA Methyl Marks

PAGE 879 and 868 and 893

Although methylation of RNA on N⁶-adenosine has attracted recent attention, three papers in this issue reveal the presence and properties of this modification in the DNA of eukaryotes, including two metazoans, *Drosophila* and *Caenorhabditis*, long thought to lack meaningful DNA methylation. Fu et al. characterize a periodic distribution of 6mA in the algae *Chlamydomonas* and find that it associates with transcription start sites of active genes. Greer et al. identify the mark in *C. elegans* along with writing and erasing enzymes and provide evidence pointing to its role in transgenerational inheritance. Finally, Zhang et al. show that 6mA in *Drosophila* DNA correlates with transposon expression and is regulated by the *Drosophila* Tet homolog, shown by the authors to be essential for development.

Hoarding Increases with Age

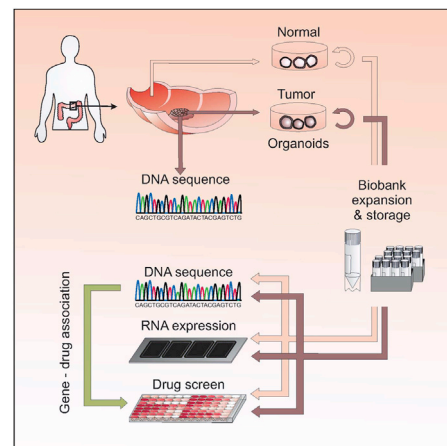
PAGE 919

Aging is associated with a decline in protein homeostasis which may affect cellular and organismal functions. A large-scale quantitative analysis of the *C. elegans* proteome along the lifespan by Walther et al. demonstrates widespread proteome imbalance and protein aggregation in aged organisms. Notably, increased formation of insoluble aggregates associated with molecular chaperones correlates with extended lifespan, suggesting that sequestering aberrant proteins delays proteostasis decline during aging.

Organic Personalized Medicine

PAGE 933

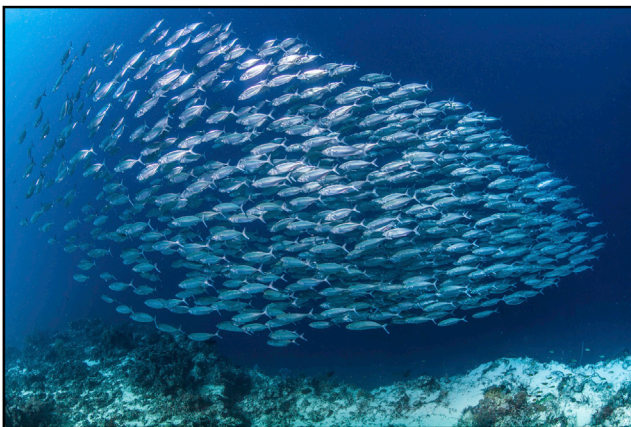
In a push toward personalized medicine, van de Wetering et al. develop three-dimensional organoids from healthy donors and patients with colorectal cancer patients. They show that tumor organoids not only closely model tumors in terms of copy number and mutation spectra but are also amenable to high-throughput drug screens, allowing for the detection of personalized gene-drug associations.



The Wisdom of Crowds

The ability of some animals to mount highly coordinated collective responses such as schooling, swarming, and flocking is extraordinary and somewhat unsettling, as anyone who watched the harmonious navigation of a school of fish can attest. Great strides have been made since the early days of studying this phenomenon, when telepathy seemed as good an explanation as any (Selous, 1931), but the question still stands: how do hundreds of minds act as one?

Early studies of group behaviors showed that they arise from individual decisions that are transmitted to the collective, and several recent papers, buoyed by technological advances that allow analysis and manipulations of complex behaviors, tackled the mechanisms of this process. Work by Richard Benton and colleagues (Ramdya et al., 2015) focused on mapping the responses of individuals that orchestrate group behaviors. For their analysis, Benton and colleagues zoomed in on *Drosophila melanogaster*, a model organism whose genetic, cellular, and circuit underpinnings are well characterized. *Drosophila* is a solitary species that doesn't typically display swarming behaviors; however, the authors found that a noxious stimulus such as CO₂ elicited a stronger avoidance response in flies that were part of a group than solitary animals. Strikingly, this response seemed to rely on communication, with escape behavior being initiated upon interactions among neighboring flies. Flies communicated the perceived danger by tapping each other with their appendages, and genetic and optogenetic manipulations mapped the circuit effectors to specific mechanosensory neurons and channels. This link between mechanosensation and group behaviors shows that more sensitive individuals can communicate, perhaps unconsciously, a stimulus to the less perceptive ones, initiating cascades of directed locomotion and coherent movement away from the stimulus.



Large school of mackerel. Image from iStock.com/paulbcowell.

While Benton and colleagues explored how individuals give rise to group behaviors, the researchers from Iain Couzin's laboratory (Rosenthal et al., 2015) turned the question on its head by studying how a complex social milieu translates

into the behavioral responses of individuals. By tracking the positions and body postures of fish in a school, they were able to reconstruct the visual information available to each individual and determine which social cues informed their decision to respond. This reverse engineering of a school's responses allowed them to identify the most influential individuals in a group and determine what characterizes likely "first responders." Thus, uncovering the communication channels among individuals can explain how seamless group decision-making happens.

Collective behaviors are commonly associated with responses to danger; however, they are also used to crowd-source intelligence for complex tasks. Studies in bats and ants demonstrated the wisdom of group intelligence, but they also mapped its limitations. By tagging them with GPS and microphones, Yossi Yovel and colleagues recorded the behaviors of bats that forage in groups and found that they eavesdrop on echolocation signals from their group mates to increase the probability of finding prey (Cvikel et al., 2015). Such "public information" is clearly useful, but it can also become a nuisance as signal interference increases with group size, impairing prey detection. Dynamics driven by group size are also at play in ant decision making, as shown by Stephen Pratt and colleagues (Sasaki et al., 2013), where positive feedback and a quorum rule among group members can direct the integration and sharpening of complex decisions during house hunting, but also lock the colony onto a suboptimal choice for less challenging tasks.

In summary, technical advances paved the way for dissecting the complexity of collective behaviors from the molecular to the behavioral level. Let's turn our collective attention to these developments, as exciting news on this front is may also inform our own, increasingly crowdsourced world.

REFERENCES

- Cvikel, N., Egert Berg, K., Levin, E., Hurme, E., Borissov, I., Boonman, A., Amichai, E., and Yovel, Y. (2015). *Curr. Biol.* 25, 206–211.
- Ramdya, P., Lichocki, P., Cruchet, S., Frisch, L., Tse, W., Floreano, D., and Benton, R. (2015). *Nature* 519, 233–236.
- Rosenthal, S.B., Twomey, C.R., Hartnett, A.T., Wu, H.S., and Couzin, I.D. (2015). *Proc. Natl. Acad. Sci. USA* 112, 4690–4695.
- Sasaki, T., Granovskiy, B., Mann, R.P., Sumpter, D.J.T., and Pratt, S.C. (2013). *Proc. Natl. Acad. Sci. USA* 110, 13769–13773.
- Selous, E. (1931). *Thought-Transference (or What?) in Birds* (London: Constable and Company).

Mirna Kvajo

In Tight Times, Companies Fill the Funding Gap

With federal budgets under pressure, scientists turn to corporations for research support.

Earlier this year, Senator Elizabeth Warren introduced a bill that she said would provide billions of new dollars for medical research. The Massachusetts Democrat proposed that if large pharmaceutical companies are caught breaking laws, any settlements they reach with the federal government should include paying into a fund that would benefit the NIH. Warren says that such a “swear jar,” as she calls her Medical Innovation Act, would have provided roughly \$6 billion a year to the NIH research budget had it been the law over the last five years.

Pharmaceutical companies oppose the measure, and it’s not clear that it will pass. But it does call attention to a problem people in the biomedical research field agree exists, a shrinking pool of government money for funding science. Though Congress doubled the NIH budget between 1998 and 2003, it’s been contracting ever since, dropping by more than 22 percent in inflation-adjusted dollars, according to the Federation of American Societies for Experimental Biology. Public funding for scientific research has drop-

ped in Europe as well, as governments imposed austerity measures in response to the 2008 fiscal crisis. “I think all countries are struggling with their budgets,” says Birgitte Nauntofte, executive director of the Novo Nordisk Foundation in Hellerup, Denmark. “I have not heard of a country that’s not struggling.”

So researchers are looking to another source of funding—corporations. Whether through sponsored research agreements, innovative ideas about investment funds focused on science, or Denmark’s tax model that allows the Novo Nordisk Foundation to support scientists, companies are picking up the tab for science that’s not being covered by public funds.

The Whitehead Institute for Biomedical Research, in Cambridge, MA, for instance, has turned to more sponsored research to cover its overall research budget, which has stayed at roughly \$60 million for the past decade, adjusting for inflation. “Ten years ago the majority of that funding came from federal sources, and today it’s no more than a third,”

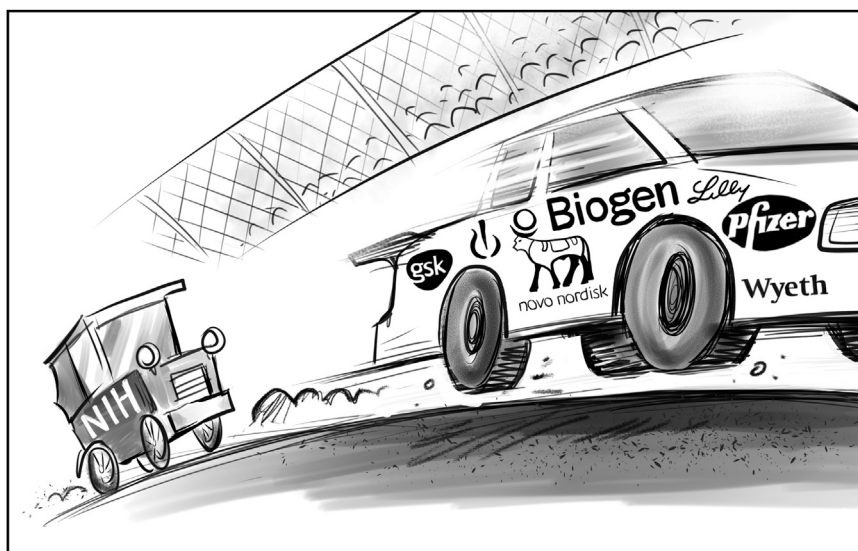
says Richard Young, a member of the institute who studies the regulatory circuitry that controls gene expression.

To make up for that drop in government grants, the Whitehead looks to other sources—philanthropy, royalties on patents, and sponsored research agreements. Last year, for instance, Whitehead announced it had signed a 3-year deal in which the biotechnology company Biogen would provide \$5.25 million to fund basic research in immunology, neurology, developmental biology, genetics, and genomics. “That’s basically an R01 level of funding,” says Mark Muskavitch, senior director of epigenetics at Biogen, making Biogen’s support for a given project comparable to the NIH’s. He runs a consortium that also includes researchers from Harvard Medical School, Brigham and Women’s Hospital, Institut Pasteur in Paris, and Washington University in St. Louis, MO, and focuses on the biology of neurodegeneration. Biogen is also funding other consortia looking at amyotrophic lateral sclerosis, scleroderma, fibrosis, and sickle cell anemia.

“Biogen has been and wants to remain an innovative company, and innovation comes from research,” Muskavitch says. These are not, he stresses, outsourced corporate research and development programs, aimed at producing products for the company. “It’s not exactly blue sky, but only secondarily are we trying to move in limited cases into translational work,” he says. “We’re trying to encourage basic research with an eye toward translatability but not a requirement for translation.”

That said, the company is putting its money into areas of biology where its markets lie, such as treatment for Alzheimer’s or Parkinson’s disease. Muskavitch says in selecting the projects to fund within the Whitehead, he has “encouraged but not constrained” the researchers to lean toward neurobiology.

To receive Biogen funding, Whitehead researchers go through a grant-writing and approval process that’s similar to applying for NIH money, though dealing with the company is simpler than dealing with the government, Young says. “I would say that relationship we have with Biogen is probably an easier, more friendly, more productive relationship than a comparable one with NIH,” he says.



Whitehead takes care to maintain independence in its scientific work. If the Institute has a substantial financial interest in a company, it won't take funding from that company. And none of the research can be secret. "We have to be able to publish what we learn. We can't have a restriction on publication," Young says. The same applies to government funding; Whitehead won't accept Department of Defense funding that comes with publication restrictions.

What Whitehead does provide to funders is an advance look at research results, often for a 30-day period, as long as that doesn't delay publication. Under some agreements, companies also have a right of first refusal over any intellectual property the Whitehead develops. Another benefit to corporations that sponsor research is developing relationships with scientists who work in research areas of interest to those companies. "Components of every major pharma are here in the Boston area because they want to be close to that human capital," Young says.

In Denmark, the Novo Nordisk Foundation also tries to use its funds to generally support basic research and to develop experts, while still having an eye on advances that could benefit its areas of specialization, which include diabetes, hemophilia, and hormone replacement therapy. "The overall goal of our grants is we would like to develop what we call a knowledge-based society," says Nauntofte. "We also want to help foster a world class educational system."

The foundation provides grants totaling 785 million Danish krone (~US \$113 million), and plans to increase that to 1.5 billion krone (~US \$216 million) by 2018. Half of that goes to support health-related science, with focuses on endocrinology and metabolic physiology. Another 20% goes to biotechnology, including finding new methods for synthesis and production. About 10% goes to education and another 10% to humanitarian purposes, even art history. And roughly 4% goes to supporting research that could lead researchers to start their own companies. The foundation has established centers for metabolic research, biosustainability, protein research, and basic stem cell biology. It also founded the Danish National Biobank to collect

biological samples from the population at large.

The Danish foundation structure is unusual, Nauntofte says. Foundations, often created by the founders of successful companies, actually own their companies and receive tax benefits for giving away a percentage of their profits. "Some of our best-performing companies are owned by private foundations and it's quite unique to our country," Nauntofte says. She estimates that approximately 80 percent of what Denmark spends on research comes from the government and about 10 percent comes from private foundations. Another 7 percent comes from European and American funding agencies, with private companies covering the other 3 percent.

The Novo Nordisk Foundation maintains a controlling interest in the publicly traded pharmaceutical company, Novo Nordisk A/S, and Novozymes, a biotechnology company that manufactures enzymes. While the public company invests in other life-science companies, the foundation controls the grant process. The process is similar to that of any funding agency, with calls for applications, deadlines, and a review by a panel of experts, none of whom can be Novo Nordisk employees. But once the award is granted, Nauntofte says, there are no strings attached. "They're all donations, they have no restrictions," she says. "The researchers have full freedom. We give the money away and it's all theirs."

Søren Molin, a systems biologist at the Technical University of Denmark, in Lyngby, receives about \$1 million of funding a year as scientific director of the bacterial cell factories section of the foundation's Center for Biosustainability. "We have the obligation in the contract that we have to produce science of the greatest impact and the greatest quality," he says. "The other obligation we have is to burn the money." The recipients have to spend the funds, not save or invest it or return it to the foundation.

The center was created in 2010 after the foundation approached the university and asked them to propose a project that could be labeled "biosustainability." Beyond that initial direction, Molin says, the foundation does not tell scientists what work to do. "We are doing biotech but not necessarily the kind of biotech

that Novo Science is doing," Molin says, though he adds, "Some of the research and technology we're doing might be useful and benefit the company eventually."

He doesn't think the center, which will receive nearly \$160 million over 10 years, would exist if it had to rely on Danish funding agencies. "There would be no way in this country that any research council could spend this kind of money toward a specific issue," he says.

In countries that don't have Denmark's tax structure, there are still creative ways to funnel corporate cash into research. Google, for instance, formed a biotechnology company, Calico, in 2013 to focus on diseases of aging, including neurodegeneration and cancer, with an initial investment of \$240 million and the promise of up to another \$490 million. Calico hired Arthur Levinson, former CEO of Genentech, to run the company and brought other highly respected scientists on board. Last year, Calico joined forces with the biopharmaceutical company AbbVie to create an R&D collaboration, with AbbVie contributing \$750 million. Calico declined requests for an interview. Meanwhile, Google's research arm, Google X, has a life sciences division, which is developing wearable health sensors and planning to collect genetic and molecular information from thousands of people. The Wall Street Journal last July reported that the life sciences division had built a team of 70 to 100 experts in areas such as physiology, biochemistry, optics, imaging, and molecular biology.

But turning to individual companies might not be the only way to find money for science. Andrew Lo, a professor of Finance at MIT's Sloan School of Management, proposes creating investment funds, not unlike the mutual funds in which people invest their IRAs. There's a gap, he says, between the basic research funded by the government, and potentially marketable therapies, supported by biopharmaceutical investors once they've made it through Phase 2 of clinical trials. He'd like to fill that gap with megafunds, large pools of investment dollars that could support research in places like the Whitehead.

A megafund, Lo explains, might pick out the top 50 or so biomedical research institutions in the country, and invest in five or

10 labs in each center. The labs could be selected by an expert review committee, and some could be culled from the fund based on progress reports. Beyond that review, the researchers would have free rein, with the stipulation that if they developed anything marketable, investors would receive 8 percent of the royalties. If such an investment produced only one or two new multi-million-dollar drugs, it would pay off handsomely. “It only takes one or two cancer drugs to generate profits, but you pay for all of the losses in a diversified portfolio,” Lo says.

In a paper published in *Nature Biotechnology* in 2012, Lo ran a simulated fund based on historical data from the previous two decades and estimated that a fund of \$5 billion to \$15 billion could generate a return of anywhere from 5 to 12 percent, depending on how it was set up. In a separate simulation, he found that a megafund taking advantage of orphan drug rules could generate a return in double digits if it invested \$575 million in from 10 to 20 projects. He points to the Cystic Fibrosis

Foundation, which gave the drug company Vertex \$150 million. Vertex developed a treatment for CF, and last year the CF foundation sold the royalties from that drug for \$3.3 billion, which it can now apply to further research.

“It could be a more sustainable way for science to become self-supporting,” says Lo, who expects that small versions of a megafund could arise within the coming year.

All this talk of corporate funding may lead to worries about privatizing science, with government leaving support of research to the private sector. Molin, for example, says that since he received money from the Novo Nordisk Foundation, it’s been harder for him to get grants from Danish research councils—he does better with European Union funders. “It’s going to be an interesting situation to see within say the next five years how the balance will be between private and public funding of research, Molin says. “If this balance is too biased in one direction, it’s not so healthy.”

Nauntofte says that most policy-makers realize that providing a stable research and educational system is the purview of federal governments. “The idea is of course not to substitute for the government, the idea of the foundation is to make supplements,” she says.

Muskavitch doesn’t believe corporate dollars can make up for lack of federal investment in research, and he worries about the fact that investment has been shrinking. “NIH has become dysfunctional, and the scientific enterprise in the US is at great risk going forward in remaining at the leading edge of biological discovery and other discovery,” he warns.

And Young argues that basic research, which not only provides new discoveries but also acts as an economic engine, needs public support. “The federal government has the responsibility of ensuring that that basic research is healthy,” Young says. “The system is really highly dependent on the government recognizing its role.”

Neil Savage

Lowell, MA

<http://dx.doi.org/10.1016/j.cell.2015.04.035>

Figuring Fact from Fiction: Unbiased Polling of Memory T Cells

Carmen Gerlach,¹ Scott M. Loughhead,¹ and Ulrich H. von Andrian^{1,2,*}

¹Harvard Medical School, Department of Microbiology and Immunobiology, Division of Immunology, Boston, MA 02115, USA

²The Ragon Institute of MGH, MIT, and Harvard, 400 Technology Square, Cambridge, MA 02139, USA

*Correspondence: uva@hms.harvard.edu

<http://dx.doi.org/10.1016/j.cell.2015.04.038>

Immunization generates several memory T cell subsets that differ in their migratory properties, anatomic distribution, and, hence, accessibility to investigation. In this issue, Steinert et al. demonstrate that what was believed to be a minor memory cell subset in peripheral tissues has been dramatically underestimated. Thus, current models of protective immunity require revision.

In 1936, *Literary Digest*, a political magazine, surveyed a quarter of the U.S. voting population and predicted that Senator Alfred Landon would capture 55% of the vote and defeat the incumbent Franklin D. Roosevelt. On Election Day, Roosevelt soundly defeated Landon with 61% of the vote, the largest margin of victory in history at the time. How could the magazine's polling have been this embarrassingly misleading? The answer lay in the methodology that was used, particularly the inherently biased sampling of respondents whose names could be easily obtained from phone directories and automobile registration records, a group that was not representative of contemporary U.S. voters (Squire, 1988). This kind of bias easily creeps into political polls, and careful measures are now being taken to avoid such pitfalls.

In science, however, we sometimes forget that the methodologies we use can similarly skew what appear to be objective outcomes. In this issue, Steinert et al. (2015) provide a telling example of how a widely used analytical approach in cellular immunology has distorted the field's concepts of immune surveillance by memory T cells. The authors demonstrate that the traditional approach relied on data extrapolation from apparently non-representative samples and the use of unreliable surrogate markers for functional definitions of cellular subsets.

Immune challenges, such as infections or vaccination, result in the activation (also called "priming") of naive T lymphocytes in secondary lymphoid organs (SLOs). Some of the activated T cells

differentiate into so-called memory cells, which have the capacity to persist for many years after the original challenge has been cleared. Importantly, memory cells provide enhanced protection against re-infection with the same pathogen. Memory T cells are usually classified into three distinct subsets based on each subset's unique migratory behavior (Mueller et al., 2013; Sallusto et al., 1999). Central memory T cells (T_{CM}) circulate through blood and SLOs, including the lymph nodes, which collect lymph fluid from the body's peripheral tissues. Effector memory T cells (T_{EM}) lack lymph node homing capacity; T_{EM} are found in blood and spleen and were widely assumed to also survey non-lymphoid tissues. More recently, a third memory T cell population was identified: the tissue resident memory T cells (T_{RM}). T_{RM} arise soon after priming from activated effector cells that seed peripheral tissues. Unlike T_{EM} , which have been thought to visit such tissues transiently, T_{RM} are largely sessile and do not circulate. Recent studies revealed that, at least in some settings, T_{RM} are more effective at protecting non-lymphoid tissues from pathogens than the migratory T_{CM} and T_{EM} (Mackay et al., 2012). This posed an apparent conundrum because T_{RM} were believed to be sparse and vastly outnumbered by their neighboring parenchymal cells. Since T cells must directly touch every infected cell that they are meant to protect, how could the rare T_{RM} be so effective at protecting the abundant somatic cells from invading pathogens?

An early glimpse of the overall distribution of the memory T cell repertoire in

immunized mice was provided in 2001 by two classical studies that showed that most memory cells reside in peripheral tissues and not in SLOs (Masopust et al., 2001; Reinhardt et al., 2001). One of these studies tracked CD4 memory cells by immunohistochemical analysis of whole-body sections of immunized mice (Reinhardt et al., 2001), a tour-de-force strategy that yields unbiased results but is technically highly demanding. Thus, more recent studies in the field have resorted to quantifying memory T cells in single-cell suspensions of tissues that were freshly harvested from immunized mice (Figure 1). To distinguish between the different memory cell subsets, researchers stain the recovered T cells with antibodies to lymph node homing receptors (expressed on T_{CM} , not T_{EM} or T_{RM}) and to two surface markers, CD69 and CD103, which were thought to be diagnostic for T_{RM} . Several studies have distinguished between extra- and intravascular memory cells by intravenously injecting an antibody to a common T cell surface moiety (e.g., CD45) coupled to a large fluorophore, such as phycoerythrin, a few minutes prior to sacrificing the animal. The injected antibody remains confined to the vessel lumen during this brief time interval, so it stains selectively the intravascular subset (Anderson et al., 2014). The extravascular T cells, which remain unstained, are composed of non-migratory T_{RM} and additional memory cells that access peripheral tissues sporadically from the blood and eventually depart via the draining lymphatics (Mackay et al., 1988). The latter have

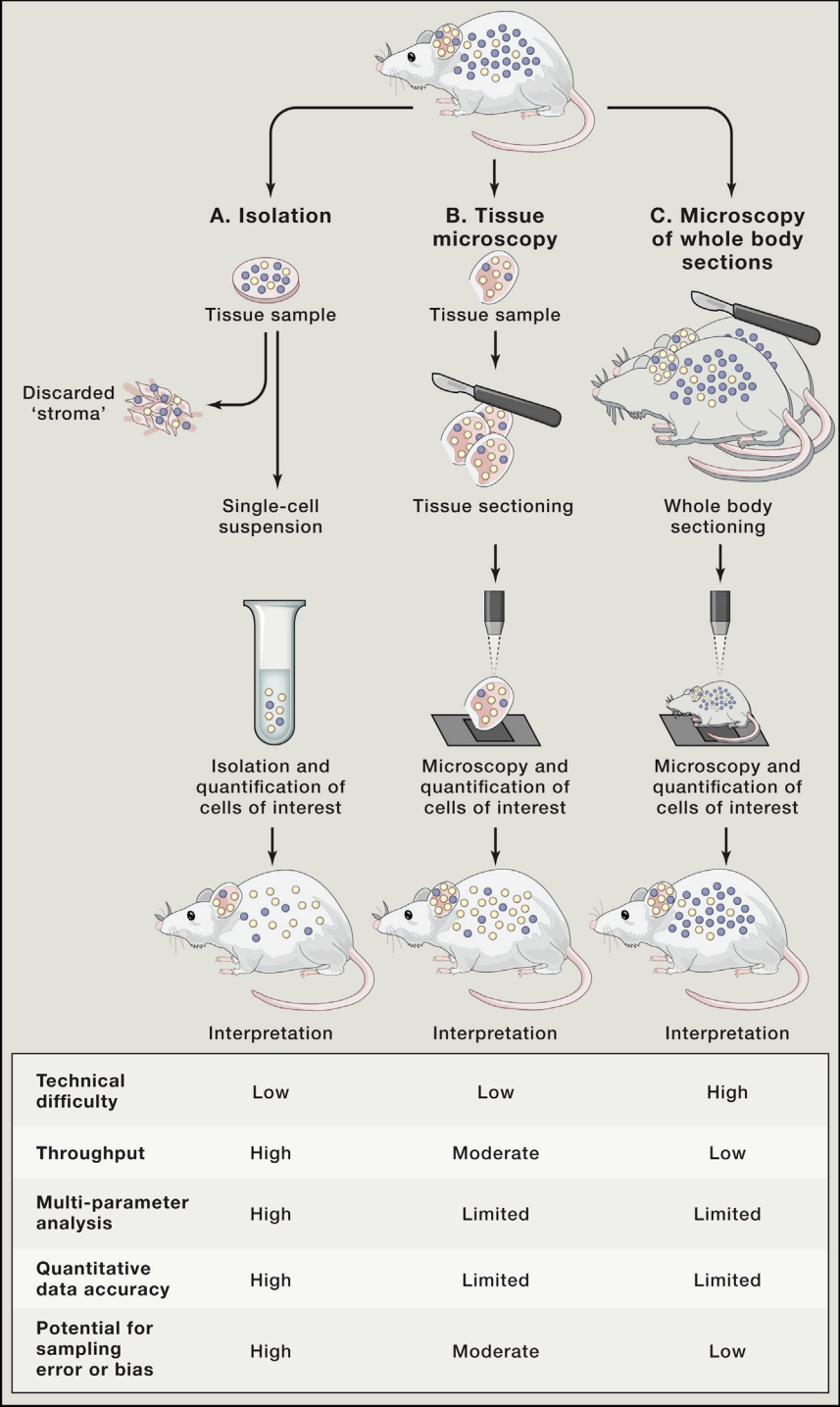


Figure 1. A Comparison of Analytical Methods to Quantify Memory T Cells in Immunized Mice

Colored spheres represent memory T cells that are non-randomly dispersed throughout the body. Different colors symbolize different subsets of memory T cells (only two subsets are shown for simplicity). (A) In the most common approach in the field, a tissue sample is enzymatically digested and mechanically dissociated to generate a single-cell suspension, while indigestible tissue stroma is discarded. In this approach, T cell isolation is often incomplete, isolation efficiency can vary between T cell subsets, and information regarding the spatial localization of the T cells within the tissue and body is not preserved. (B) Tissue samples are sectioned and analyzed by immunostaining and quantitative microscopy. Data

long been assumed to be recruited from the T_{EM} subset, although experimental evidence has largely been lacking.

These standard procedures for memory cell isolation have been relying on two assumptions: (1) T cell isolation from tissue-derived cell suspensions is efficient and yields every memory subset without bias, and (2) the presence and identity of T_{RM} is faithfully reported by CD69 and/or CD103 expression combined with lack of intravascular staining. In this issue, Steinert et al. test both assumptions by comparing the frequency and phenotype of each memory subset recovered from traditional tissue suspensions with results obtained using exacting quantitative microscopy of immunostained tissue sections (Figure 1). The results are unexpected. The number of T_{RM} that are found in sections of some peripheral tissues, such as the female reproductive tract (FRT), is much larger (by as much as 60-fold) than the number of T_{RM} that can be recovered from single-cell preparations of the same tissues. This discrepancy reflects a dramatic loss of T cells during tissue processing, whereby many cells are presumably either killed or discarded with indigestible tissue stroma. T cell loss disproportionately affects the recovery of T_{RM} , resulting in over-representation of other memory subsets, particularly those in the intravascular compartment. Furthermore, when the two analytical techniques are applied to other tissues, such as spleen and lymph nodes, both approaches yield comparable numbers of memory cells. These findings imply that the standard model of peripheral T cell memory, which has been largely based on analyses of tissue suspensions, not only underestimates the overall size of the memory pool, but also is based on a severely skewed perception of subset abundance both between different

from the analyzed region/tissue is extrapolated to the whole organ and even the whole mouse. Information regarding the density and spatial distribution of T cell subsets within the analyzed sample is well conserved, but results may not necessarily be representative of the whole mouse. (C) Analysis of whole-body sections by microscopy can provide information regarding the spatial distribution of T cell subsets within an entire animal; however, the approach is technically very demanding.

anatomic regions and within any given tissue.

Steinert et al. also interrogate the second assumption: that T_{RM} faithfully express CD69 and/or CD103 and are not accessible to intravascular antibody. Using parabiotic pairs of congenic mice, which were surgically joined to establish a shared blood circulation, the authors discover that a sizeable fraction of T_{RM} express neither CD69 nor CD103, and some T_{RM} , especially in the kidney and liver, actually appear to reside within the intravascular space.

These findings have implications for how immunologists think about T cell surveillance of tissues, particularly with regard to T_{RM} . For example, in the FRT, isolation-based methods had estimated that there is one T_{RM} for every ~20,000 nucleated cells, while tissue microscopy performed by Steinert et al. reveals that there is one T_{RM} for every ~300 nucleated cells. Assuming that T_{RM} within the FRT scan cells at a similar rate to those in the skin (Ariotti et al., 2012), isolation-based methods project that T_{RM} would require

~1 month to scan every cell in the FRT. In contrast, the tissue microscopy data imply that T_{RM} scan the FRT in its entirety within ~12 hr, an estimate that is much more consistent with the reported effectiveness of T_{RM} to protect non-lymphoid tissues (Mackay et al., 2012).

Steinert and colleagues thus provide a much-needed reality check for immunologists. Their findings will have to be taken into account when evaluating immune responses to vaccines and pathogens, and it will be important to determine their impact on our understanding of allergic and autoimmune diseases, as well as immuno-oncology.

Even though 80 years have passed since the *Literary Digest* fiasco, this study provides a stern reminder that sample bias is not a fiction of the past but remains to this day a fact to be reckoned with—by scientists and voters alike.

REFERENCES

Anderson, K.G., Mayer-Barber, K., Sung, H., Beura, L., James, B.R., Taylor, J.J., Qunaj, L., Grif-

fith, T.S., Vezys, V., Barber, D.L., and Masopust, D. (2014). *Nat. Protoc.* 9, 209–222.

Ariotti, S., Beltman, J.B., Chodaczek, G., Hoekstra, M.E., van Beek, A.E., Gomez-Eerland, R., Ritsma, L., van Rhee, J., Marée, A.F., Zal, T., et al. (2012). *Proc. Natl. Acad. Sci. USA* 109, 19739–19744.

Mackay, C.R., Kimpton, W.G., Brandon, M.R., and Cahill, R.N. (1988). *J. Exp. Med.* 167, 1755–1765.

Mackay, L.K., Stock, A.T., Ma, J.Z., Jones, C.M., Kent, S.J., Mueller, S.N., Heath, W.R., Carbone, F.R., and Gebhardt, T. (2012). *Proc. Natl. Acad. Sci. USA* 109, 7037–7042.

Masopust, D., Vezys, V., Marzo, A.L., and Lefrançois, L. (2001). *Science* 291, 2413–2417.

Mueller, S.N., Gebhardt, T., Carbone, F.R., and Heath, W.R. (2013). *Annu. Rev. Immunol.* 31, 137–161.

Reinhardt, R.L., Khoruts, A., Merica, R., Zell, T., and Jenkins, M.K. (2001). *Nature* 410, 101–105.

Sallusto, F., Lenig, D., Förster, R., Lipp, M., and Lanzavecchia, A. (1999). *Nature* 401, 708–712.

Squire, P. (1988). *Public Opin. Q.* 52, 125–133.

Steinert, E.M., Schenkel, J.M., Fraser, K.A., Beura, L.K., Manlove, L.S., Igyártó, B.Z., Southern, P.J., and Masopust, D. (2015). *Cell* 161, this issue, 737–749.

Shedding Light on Glioma Growth

Emily K. Lehrman¹ and Beth Stevens^{1,*}

¹Department of Neurology, F.M. Kirby Neurobiology Center, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA

*Correspondence: beth.stevens@childrens.harvard.edu

<http://dx.doi.org/10.1016/j.cell.2015.04.036>

Cancer is known for opportunistically utilizing resources from its surroundings for its own growth and survival. In this issue of *Cell*, Venkatesh et al. demonstrate that this also occurs in the brain, identifying neuronal activity-induced secretion of neuroligin-3 as a novel mechanism promoting glioma proliferation.

Cancer is notorious for hijacking normal biological processes to promote tumor cell survival, migration, and proliferation. Cancer cells release angiogenic factors that promote blood vessel formation to support their own survival and upregulate molecules normally expressed by healthy cells to evade immune detection. In their recent study, Venkatesh et al. (2015) reveal that cancer cells also take advantage of neuronal activity, the most essen-

tial aspect of brain function, in order to proliferate. The authors demonstrate that optogenetic stimulation of neurons can promote the growth of human high-grade gliomas (HGGs) by inducing the secretion of mitogenic factors.

This study was initiated following the discovery that neuronal activity stimulates the proliferation of oligodendrocyte precursor cells (OPCs) and neuronal precursor cells (NPCs) in vivo (Gibson et al.,

2014), cells that can give rise to gliomas (Cuddapah et al., 2014). Both studies utilized optogenetic strategies to increase neuronal activity by stimulating channelrhodopsin-expressing neurons with blue light (Figure 1A). This approach enables the activation of subsets of neurons in defined circuits in a physiological manner and allows for comparisons between different circuits or regions from within the same brain. Importantly, this method

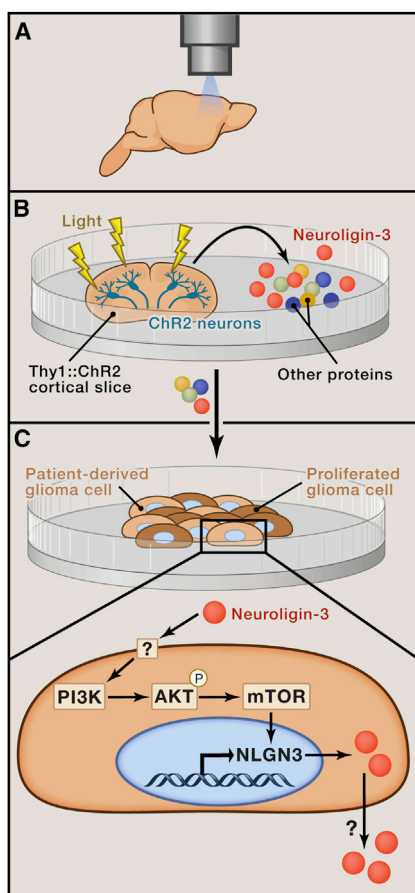


Figure 1. Optogenetic Techniques Reveal that Activity-Induced Secretion of Neuroligin-3 Promotes Glioma Growth

(A) In vivo optogenetic stimulation of Thy1::ChR2 premotor cortex promotes the proliferation of xenografted glioblastoma cells.

(B) In vitro optogenetic stimulation of Thy1::ChR2 cortical slices leads to activity-regulated secretion of factors into the media.

(C) Conditioned medium from stimulated cortical slices induces growth of patient-derived glioma cells in vitro. Venkatesh et al. identify secreted neuroligin-3 as their primary candidate mitogen and propose a downstream signaling pathway involving PI3K and mTOR.

can be used to stimulate the brains of awake, behaving animals.

Venkatesh et al. put this technique to use in their orthotopic xenograft model of pediatric HGG. To create this model, the authors xenografted cells cultured from a biopsy of frontal cortex glioblastoma from a 15-year-old patient into the premotor cortex of immunodeficient mice that had been crossed to the Thy1::ChR2 line, which would allow for activation of neural tissue surrounding the xenografted cells. Just as in non-pathological experiments

with OPCs and NPCs, optogenetic stimulation of neuronal activity using blue light induced proliferation of xenografted pediatric HGG cells. A single stimulation was sufficient to induce proliferation; however, repetitive stimulation over the course of a week further increased tumor cell burden.

The authors next moved to an in vitro system to investigate the mechanism underlying neuronal activity-induced glioma proliferation (Figure 1B). After determining that conditioned medium collected from optogenetically stimulated Thy1::ChR2 cortical slices could induce the proliferation of a variety of different patient-derived HGG cell cultures (Figure 1C), Venkatesh et al. sought to identify the secreted signal responsible. Glioma cells express ion channels and neurotransmitter receptors and are sensitive to calcium, and, therefore, could proliferate in response to a variety of secreted signals (Cuddapah et al., 2014). Using mass spectrometry, the authors identified secreted neuroligin-3 in the cortical slice-conditioned medium as their primary candidate mitogen and confirmed its ability to induce the proliferation of multiple types of HGG using a recombinant protein. Importantly, the neuroligin-3 found in the conditioned medium contained only the ectodomain, suggesting that it is cleaved in a similar manner to known family member neuroligin-1 (Peixoto et al., 2012; Suzuki et al., 2012).

To understand how neuroligin-3 could exert this mitogenic effect, the authors performed RNA sequencing followed by western blot analysis on cultured glioma cells that had been treated with either light-exposed WT or Thy1::ChR2-conditioned medium. They determined that neuronal activity-regulated secretion of neuroligin-3 promoted glioma cell proliferation through activation of the PI3K-mTOR pathway (Figure 1C). Interestingly, this pathway activated both transcription and translation of neuroligin-3 in glioma cells, suggesting a feedforward signaling loop. Increased neuroligin-3 expression by tumor cells may indeed be pathogenic, as the authors found an inverse relationship between adult glioblastoma neuroligin-3 mRNA expression and patient survival upon analyzing data from The Cancer Genome Atlas. On average,

patients categorized as having high levels of glioblastoma neuroligin-3 expression had a lifespan that was 5 months shorter than that of patients with low expression of neuroligin-3. This effect was specific, as there was no association between expression of neuroligin-2, which does not induce glioma cell proliferation in vitro, and patient survival.

The finding that neuronal activity promotes glioma proliferation raises a number of interesting questions. Previous studies have found that cancer cells secrete glutamate, which may stimulate their own proliferation, and that glutamate secretion may also be linked to the epilepsy developed by many glioma patients (Buckingham et al., 2011). Given this new link between neuronal activity and glioma growth, one question that arises is whether tumor-associated epilepsy serves as another type of feedforward mechanism to support further glioma proliferation. Additionally, this link may also shed light on studies that demonstrate an increased likelihood of brain tumor development in patients who have been treated for epilepsy, which is not very well understood (Khan et al., 2011).

The discovery of neuroligin-3 as a potential mitogen is also unexpected, given its role as a postsynaptic adhesion molecule required for normal synaptic function and its implication as a disease gene in autism (Südhof, 2008). This study suggests that neuroligin-3 may play other roles in non-neuronal cells and that secreted neuroligin-3 may regulate its own transcription. Additional work is necessary to understand the mechanism outlined in this study, including the cellular origin of secreted neuroligin-3, the activity-dependent neuroligin-3 cleavage mechanism, the recruitment of downstream PI3K, and the function of tumor-derived neuroligin-3. Some insight into neuroligin-3 cleavage may be gleaned from studies that have examined neuroligin-1, which has been shown to exhibit activity-dependent cleavage and secretion of its ectodomain (Peixoto et al., 2012; Suzuki et al., 2012). These findings also highlight the need to investigate the role of secreted neuroligin-3 in the healthy brain and whether the feedforward pathway is utilized for growth of healthy OPCs and NPCs or only arises in glioma cells.

Venkatesh et al. provide invaluable insight into HGG, revealing not only a greater mechanistic understanding of the regulation of glioma growth, but also a potential therapeutic target in neuroligin-3. Their observations that neuronal activity promotes the proliferation of multiple glioma types and that neuroligin-3 is mutated in a variety of different types of cancers, combined with recent studies implicating autonomic innervation with cancer progression in other systems (Magnon et al., 2013; Zhao et al., 2014), suggest that this mechanism may be broadly applicable to many cancers.

REFERENCES

- Buckingham, S.C., Campbell, S.L., Haas, B.R., Montana, V., Robel, S., Ogunrinu, T., and Sontheimer, H. (2011). *Nat. Med.* 17, 1269–1274.
- Cuddapah, V.A., Robel, S., Watkins, S., and Sontheimer, H. (2014). *Nat. Rev. Neurosci.* 15, 455–465.
- Gibson, E.M., Purger, D., Mount, C.W., Goldstein, A.K., Lin, G.L., Wood, L.S., Inema, I., Miller, S.E., Bieri, G., Zuchero, J.B., et al. (2014). *Science* 344, 1252304.
- Khan, T., Akhtar, W., Wotton, C.J., Hart, Y., Turner, M.R., and Goldacre, M.J. (2011). *J. Neurol. Neurosurg. Psychiatry* 82, 1041–1045.
- Magnon, C., Hall, S.J., Lin, J., Xue, X., Gerber, L., Freedland, S.J., and Frenette, P.S. (2013). *Science* 341, 1236361.
- Peixoto, R.T., Kunz, P.A., Kwon, H., Mabb, A.M., Sabatini, B.L., Philpot, B.D., and Ehlers, M.D. (2012). *Neuron* 76, 396–409.
- Südhof, T.C. (2008). *Nature* 455, 903–911.
- Suzuki, K., Hayashi, Y., Nakahara, S., Kumazaki, H., Prox, J., Horiuchi, K., Zeng, M., Tanimura, S., Nishiyama, Y., Osawa, S., et al. (2012). *Neuron* 76, 410–422.
- Venkatesh, H.S., Johung, T.B., Caretti, V., Noll, A., Tang, Y., Nagaraja, S., Gibson, E.M., Mount, C.W., Polepalli, J., Mitra, S.S., et al. (2015). *Cell* 161, this issue, 803–816.
- Zhao, C.-M., Hayakawa, Y., Kodama, Y., Muthupalani, S., Westphalen, C.B., Andersen, G.T., Flatberg, A., Johannessen, H., Friedman, R.A., Renz, B.W., et al. (2014). *Sci. Transl. Med.* 6, 250ra115.

Rods Feed Cones to Keep them Alive

Jacek Krol^{1,*} and Botond Roska^{1,2,*}

¹Neural Circuit Laboratories, Friedrich Miescher Institute for Biomedical Research, 4058 Basel, Switzerland

²Faculty of Medicine, University of Basel, 4056 Basel, Switzerland

*Correspondence: jacek.krol@fmi.ch (J.K.), botond.roska@fmi.ch (B.R.)

<http://dx.doi.org/10.1016/j.cell.2015.04.031>

Cone photoreceptors, responsible for high-resolution and color vision, progressively degenerate following the death of rod photoreceptors in the blinding disease retinitis pigmentosa. Aït-Ali et al. describe a molecular mechanism by which RdCVF, a factor normally released by rods, controls glucose entry into cones, enhancing their survival.

The retina is a highly sophisticated biological computer that captures an image with its photoreceptors and extracts different visual features to describe the visual scene to higher brain centers in simple and compact terms. Although photoreceptors, the rods and cones, are only two out of the sixty retinal cell types, they are exceptionally important: all image-forming vision depends on their proper function. Despite the fact that rods outnumber cones 20 to 1, human vision is mostly based on cones. Rods are distributed at the periphery of the retina and are the photosensors for low light levels. Cones are concentrated in the center of the retina and work at higher light levels. Since cones are necessary for the high-resolution color vision that enables us to read, recognize faces, and enjoy the colorful world, in the modern world

we surround ourselves with enough light to turn on the cones. Most of us spend little time in conditions where photons are scarce and, therefore, our dependence on rod function is minor. A study presented in this issue of *Cell* offers key insight into the interdependence of rods and cones, and how it is disrupted in the genetic disorder retinitis pigmentosa (Aït-Ali et al., 2015)

The genes involved in retinitis pigmentosa are primarily expressed only in rods and are important for their function (Hartong et al., 2006). If humans rely mostly on cone vision, why is this disease so severe? The reason stems from the fact that rods and cones are dependent on each other. When rods are dysfunctional but alive, as in another genetic disease called stationary night blindness, cones are functional. Indeed, patients with station-

ary night blindness are capable of living an almost normal life. However, when rods die, as happens in retinitis pigmentosa, cones sense this loss and react to it. This reaction is devastating. First, cones lose their outer segments, which serve as light detectors, causing patients to become blind. Second, on a longer timescale, the other parts of the cones progressively degenerate.

Due to the importance of cones for human vision, and their dependence on rods, two fundamental questions in retinitis pigmentosa research are why and how do cones react to rod death and how can we prevent cones from degenerating? There have been several important insights in recent years. One of these insights, originating in José-Alain Sahel's laboratory, came from the logic that if rods are necessary for cone

survival, rods may release a factor that enhances cone survival (Mohand-Said et al., 1998). Indeed, such a molecule, named rod-derived cone viability factor (RdCVF) has been identified by Thierry L  veillard and Jos  -Alain Sahel (L  veillard et al., 2004). It has been shown that, after rods die, the resulting loss of RdCVF production contributes to cone degeneration, and that externally supplied RdCVF slows down this process (Byrne et al., 2015; L  veillard and Sahel, 2010). However, the RdCVF receptor in cones has been unknown, and the mode of action for protecting cones remained unclear.

The present study by the L  veillard group (A  t-Ali et al., 2015) identify an RdCVF receptor, Basigin-1, and propose a mechanism, namely an increase in glucose transport via GLUT1 and a concomitant increase in aerobic glycolysis, that could be responsible for the protection of cones (Figure 1). The authors identify and verify Basigin-1 as the receptor of RdCVF for its trophic function in cones using numerous experimental approaches both in vitro and in vivo. After identifying the receptor, A  t-Ali et al. search for the mechanism leading to enhanced cone survival. Using co-immunoprecipitation followed by mass spectrometry and fluorescence resonance energy transfer assay, they find a glucose transporter, GLUT1, which interacts with Basigin-1. Both Basigin-1 and GLUT1 are expressed in photoreceptor inner segments and are essential for increased cone survival mediated by ectopic RdCVF administration. A  t-Ali et al. point out that cones are highly sensitive to glucose deprivation, suggesting that a glucose uptake-related pathway may underlie the ability of RdCVF to preserve cones. Consistently, using a non-metabolized glucose analog, the authors showed that exposure to RdCVF increased glucose entry into cones. Depletion of Basigin-1 and GLUT1 significantly

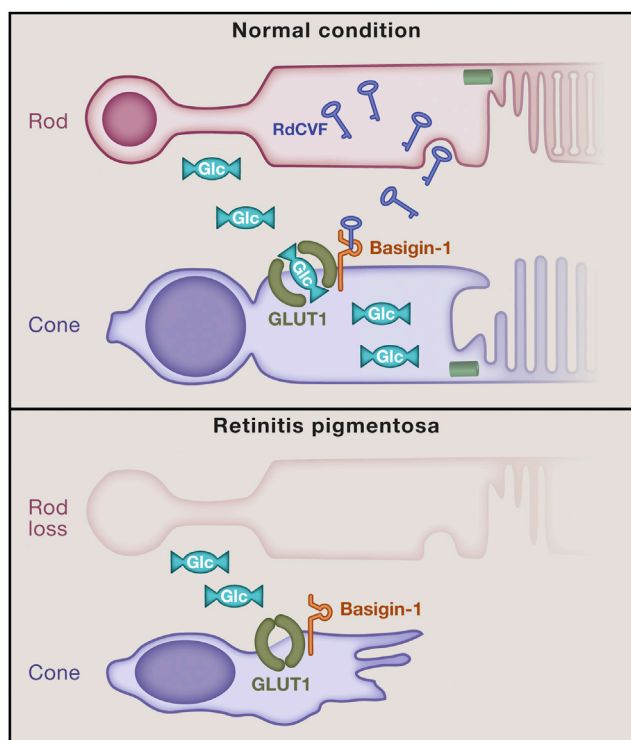


Figure 1. Rods Regulate Glucose Entry into Cones

In normal retinas, rod photoreceptors secrete rod-derived cone viability factor (RdCVF), which is necessary for cone photoreceptor survival. RdCVF binds to Basigin-1, which through the glucose transporter GLUT1, regulates glucose uptake by cones. When rods are lost, the resulting lack of RdCVF leads to cone starvation, which in turn leads to cone degeneration.

impairs RdCVF-mediated glucose uptake. How does glucose supply improve cone survival? A  t-Ali et al. observe that cones exposed to RdCVF have increased intracellular ATP concentrations and propose that ATP is produced in an unusual form of aerobic glycolysis, in which glucose is converted to lactate in the presence of oxygen. This metabolic process requires lactate dehydrogenase activity, and its inhibition abolishes RdCVF-mediated cone survival.

It has recently been shown that activation of mTORC1 increases cone survival partly by increasing glucose uptake (Venkatesh et al., 2015), suggesting that accelerating glucose entry into the cell is a convergence point for different pathways, such as RdCVF and mTOR, which protect cones. Thus, starvation appears to be a major contributor to cone degeneration in retinitis pigmentosa (Punzo et al., 2009), and feeding cones emerges as a central theme to assist in protecting them.

One of the most important implications of the identification of the RdCVF receptor Basigin-1 and its binding partner GLUT1 is the potential for developing small molecules that could activate them and, as a consequence, slow down cone degeneration in patients. One may wonder why researchers are focused on protecting cones, and not on preventing the death of rods? There are a number of reasons. First, since lack of function in rods causes few symptoms, patients often visit ophthalmologists when cones start to be affected. By this time, however, many of the rods have already degenerated. Second, rods should start to be protected before the disease starts. However, the onset of the disease, even if the affected members of a family can be determined early, is often not tractable, complicating the design of clinical trials. Despite these problems, promising new ways of protecting both rods and cones are emerging (Byrne et al., 2015).

In summary, together with exciting new gene therapy approaches to impact oxidative stress (Xiong et al., 2015), histone deacetylases (Chen and Cepko, 2009), and RdCVF (Byrne et al., 2015; L  veillard and Sahel, 2010), small molecules targeting Basigin-1 or GLUT1 may provide ways of slowing down a devastating cause of blindness.

REFERENCES

- A  t-Ali, N., Fridlich, R., Millet-Puel, G., Cl  rin, E., Delalande, F., Jaillard, C., Blond, F., Perrocheau, L., Reichman, S., Byrne, L.C., et al. (2015). *Cell* 161, this issue, 817–832.
- Byrne, L.C., Dalkara, D., Luna, G., Fisher, S.K., Cl  rin, E., Sahel, J.-A., L  veillard, T., and Flannery, J.G. (2015). *J. Clin. Invest.* 125, 105–116.
- Chen, B., and Cepko, C.L. (2009). *Science* 323, 256–259.
- Hartong, D.T., Berson, E.L., and Dryja, T.P. (2006). *Lancet* 368, 1795–1809.

Léveillard, T., and Sahel, J.-A. (2010). *Sci. Transl. Med.* 2, 26ps16.

Léveillard, T., Mohand-Saïd, S., Lorentz, O., Hicks, D., Fintz, A.-C., Clérin, E., Simonutti, M., Forster, V., Cavusoglu, N., Chalmel, F., et al. (2004). *Nat. Genet.* 36, 755–759.

Mohand-Saïd, S., Deudon-Combe, A., Hicks, D., Simonutti, M., Forster, V., Fintz, A.C., Léveillard, T., Dreyfus, H., and Sahel, J.A. (1998). *Proc. Natl. Acad. Sci. USA* 95, 8357–8362.

Punzo, C., Kornacker, K., and Cepko, C.L. (2009). *Nat. Neurosci.* 12, 44–52.

Venkatesh, A., Ma, S., Le, Y.Z., Hall, M.N., Rüegg, M.A., and Punzo, C. (2015). *J. Clin. Invest.* 125, 1446–1458.

Xiong, W., MacColl Garfinkel, A.E., Li, Y., Benowitz, L.I., and Cepko, C.L. (2015). *J. Clin. Invest.* 125, 1433–1445.

Three Cell Fusions during Double Fertilization

Stefanie Sprunck¹ and Thomas Dresselhaus^{1,*}

¹Cell Biology and Plant Biochemistry, Biochemie-Zentrum Regensburg, University of Regensburg, 93053 Regensburg, Germany

*Correspondence: thomas.dresselhaus@ur.de

<http://dx.doi.org/10.1016/j.cell.2015.04.032>

Fertilization of both egg and central cell is a major distinguishing feature of flowering plants. Now, Maruyama et al. report a third cell fusion event between the persistent synergid and the fertilized central cell shortly after double fertilization in *Arabidopsis*. This causes rapid dilution of pollen tube attractant(s), preventing polytubey.

Almost 120 years ago, Sergei Gavrilovich Navashin (1898) and Léon Guignard (1899) described independently for the first time that two fertilization events occur in lily, a major model plant at that time. The universality of this observation was confirmed in numerous flowering plant species (angiosperms) and is now widely considered as a major feature distinguishing angiosperms from all other organisms. During the double-fertilization event, one sperm cell fuses with the egg cell, forming the embryo, and a second sperm cell fertilizes the central cell, which develops into the endosperm. This sounds simple, but fertilization in angiosperms is a very complex process: the two genetically identical and immobile sperm cells are transported via the pollen tube over long distances (e.g., up to 30 cm in maize) through the maternal tissues of the flower in order to deliver them to the ovule. Many hurdles have to be taken before the pollen tube finally arrives at the embryo sac harboring the two female gametes, egg and central cell, as well as a number of accessory cells, including two synergids (Figure 1). The synergids are known as gland cells playing a leading role in pollen tube attraction and sperm release (for review, see Dresselhaus and Franklin-Tong, 2013). In species like the model plant

Arabidopsis, usually only one pollen tube arrives at the embryo sac and communicates with the synergids until the tube tip bursts simultaneously with the first synergid, termed receptive synergid. A block to polytubey (arrival of excess pollen tubes) is established soon after fertilization and minimizes the risk of polyspermy (fusion of a female gamete with multiple sperms). However, plants are capable of attracting multiple pollen tubes—for example, when gamete fusion fails—to maximize reproductive success. In *Arabidopsis*, the second synergid, persistent synergid, was shown to be responsible for polytubey in the case of fertilization failure and continues to attract pollen tubes until it degenerates (Beale et al., 2012; Kasahara et al., 2012). It was further indicated that successful double fertilization induces a block to polytubey and thus avoids the delivery of additional sperm cells to the embryo sac. But how is this block to polytubey established? Nature has an astonishingly simple solution for this problem, which is now reported in this issue of *Cell* by Maruyama et al. (2015): the persistent synergid fuses with the huge fertilized central cell (about 20 times larger volume, which even quickly increases after fertilization), and thereby pollen tube attractants are rapidly diluted. This peculiar

phenomenon was named as synergid-endosperm fusion (SE fusion; Figure 1).

Using various fluorescent markers to label the cytosol, mitochondria, and endoplasmic reticulum, Maruyama et al. show by time-lapse imaging the mixing of persistent synergid and endosperm cytoplasm about 5 hr after fertilization, when the fertilized central cell or primary endosperm nucleus starts to divide. They further show fusion of plasma membranes of both cells, which was never observed in unfertilized ovules. Even more important and significant are the experiments in which they investigate the quick dilution of the pollen tube attractant AtLURE1 (Takeuchi and Higashiyama, 2012). AtLURE1 signals quickly decrease in the degenerated receptive synergid after sperm release but remain high in the persistent synergid. A rapid decrease of AtLURE1 signals was observed to coincide with the measurements of the dilution of cytoplasmic components. The attractant disappears almost completely within 36 min after initiation of SE fusion, a time point when the primary endosperm nucleus divides. The induction of SE fusion and thus rapid dilution of AtLURE1 into the early developing endosperm is sensed by fertilization success of the central cell, but not by

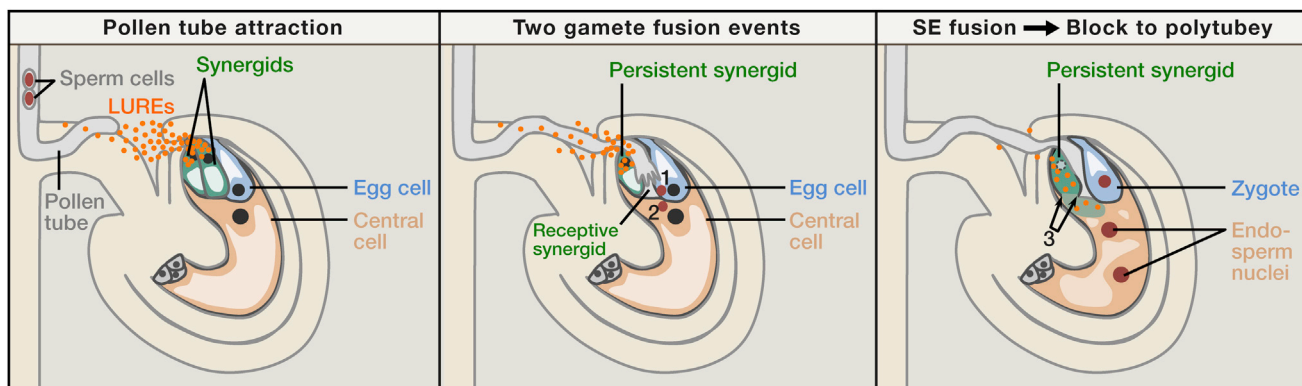


Figure 1. Three Cell Fusion Events Occur during Double Fertilization in *Arabidopsis thaliana*

(Left) The pollen tube is attracted and guided to grow into the ovule by small cysteine-rich proteins (LUREs), which are secreted by two synergids. These are part of the embryo sac, comprising additionally the two female gametes, egg and central cell, and three antipodal cells. (Middle) After pollen tube burst, two immotile sperm cells are released into the receptive synergid that degenerates. Thereafter, one sperm cell fuses with the egg cell (1) and the second sperm with the central cell (2). (Right) Maruyama et al. now show that a third cell fusion takes place after successful fertilization between the persistent synergid and the endosperm, the large fertilized central cell (3). The cytoplasm of both cells are mixed, and thus the pollen tube attractants quickly diluted. Additionally, disintegration of the persistent synergid nucleus is induced, thereby establishing a block to the attraction of excess pollen tubes.

the egg cell. However, the fertilized egg cell contributes independently to a block of polytubey as it induces rapid disintegration of the persistent synergid nucleus via an unknown ethylene response pathway and thereby additionally attenuates AtLURE1 production. In conclusion, Maruyama et al. use cutting-edge microscopic imaging of the double-fertilization process to show the establishment of a block to polytubey by two independent mechanisms (SE fusion and induced synergid nucleus disintegration), thereby eliminating the persistent synergid and its function(s) in the model plant *Arabidopsis*.

The timing of this block (~5 hr after fertilization and ~10 hr after pollination) appears late but is sufficient in *Arabidopsis*, as the arrival of secondary pollen tubes was reported to occur ~16 hr after pollination (Kasahara et al., 2012). However, this observation indicates that additional mechanisms exist, such as the direct or induced release of repellents by the first leading pollen tube, that require degradation or removal until secondary pollen tubes are attracted. Moreover, it will now be important to determine the extent to which generalizations can be made. Absorption of the synergids by the developing endosperm was also reported, for example, in

Capsella bursa-pastoris (Schulz and Jensen, 1968), indicating that this mechanism exists in all Brassicaceae species. But is this finding also relevant for other plant families, such as the economically important Gramineae (grasses)? In grasses, multiple pollen tubes arrive almost simultaneously in the vicinity of the embryo sac (Lausser et al., 2010); therefore, a polytubey block has to be established within seconds or minutes to repel excess pollen tubes. This quick reaction cannot be achieved by SE fusion or by synergid disintegration. Moreover, in plant species with more than two synergids, such as the extant most basal angiosperm *Amborella trichopoda* (Friedman and Ryerson, 2009), it remains to be investigated whether persistent synergids are all eliminated by the same mechanism. Nevertheless, SE fusion may also exist in plant species lacking a block to polytubey to remove excess synergid cells—for example, to provide more space in the embryo sac for the developing embryo. The elimination of cells is not novel, as it is a prerequisite in many tissues and organs during animal and plant development as well as during reproduction. However, the excitement about the present study is the observation that a highly active signaling cell is consumed by a much larger cell,

thereby quickly reducing the amount of secreted signaling molecules and the terminating function(s) of the absorbed cell. It will now be interesting to find this connection in other cell fusions. In conclusion, this study has demonstrated once again how simple and unexpected nature sometimes solves biological problems.

REFERENCES

- Beale, K.M., Leydon, A.R., and Johnson, M.A. (2012). *Curr. Biol.* 22, 1090–1094.
- Dresselhaus, T., and Franklin-Tong, N. (2013). *Mol. Plant* 6, 1018–1036.
- Friedman, W.E., and Ryerson, K.C. (2009). *Am. J. Bot.* 96, 129–143.
- Guignard, M.L. (1899). *Rev. Gen. Bot.* 11, 129–135.
- Kasahara, R.D., Maruyama, D., Hamamura, Y., Sakakibara, T., Twell, D., and Higashiyama, T. (2012). *Curr. Biol.* 22, 1084–1089.
- Lausser, A., Kliwer, I., Srilunchang, K.O., and Dresselhaus, T. (2010). *J. Exp. Bot.* 61, 673–682.
- Maruyama, D., Völz, R., Takeuchi, H., Mori, T., Igawa, T., Kurihara, D., Kawashima, T., Ueda, M., Itoh, M., Umeda, M., et al. (2015). *Cell* 161, this issue, 907–918.
- Navashin, S.G. (1898). *Bul. Acad. Imp. Sci. St Petersburg* 9, 377–382.
- Schulz, S.R., and Jensen, W.A. (1968). *Am. J. Bot.* 55, 541–552.
- Takeuchi, H., and Higashiyama, T. (2012). *PLoS Biol.* 10, e1001449.

An Adenine Code for DNA: A Second Life for N6-Methyladenine

Holger Heyn¹ and Manel Esteller^{1,2,3,*}

¹Cancer Epigenetics and Biology Program (PEBC), Bellvitge Biomedical Research Institute (IDIBELL), 08908 L'Hospitalet de Llobregat, Barcelona, Catalonia, Spain

²Department of Physiological Sciences II, School of Medicine, University of Barcelona, 08036 Barcelona, Catalonia, Spain

³Institució Catalana de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Catalonia, Spain

*Correspondence: mesteller@idibell.cat

<http://dx.doi.org/10.1016/j.cell.2015.04.021>

DNA N6-methyladenine (6mA) protects against restriction enzymes in bacteria. However, isolated reports have suggested additional activities and its presence in other organisms, such as unicellular eukaryotes. New data now find that 6mA may have a gene regulatory function in green alga, worm, and fly, suggesting m6A as a potential “epigenetic” mark.

The Origins of Adenine Methylation

Genetic constraints hamper the response of cells to the changing environment and represent a hurdle to adaptations that characterize living organisms. Thus, dynamic modifications that expand the genetic code beyond A, G, C, and T are necessary. Among the most studied, 5-methylcytosine (5mC) exerts a predominant role due to its important activities in mammals to establish the epigenetic setting and its relevance in human disorders, particularly cancer (Heyn and Esteller, 2012). 5mC has been named the fifth base of DNA, and only lately has a second modification in DNA, 5-hydroxymethylcytosine (5hmC), emerged as a contender for human cells (Kohli and Zhang, 2013). Other derivatives, such as 5-formylcytosine and 5-carboxylcytosine, are so far considered transitory byproducts of oxidative demethylation (Kohli and Zhang, 2013). However, this can be an anthropocentric view. N4-methylcytosine (4mC) is very common in bacteria but absent in mammals. There is an even more intriguing DNA modification: N6-methyladenine (6mA) (Figure 1A).

6mA represents a dominant modification in bacteria, while 5mC is absent in many prokaryotic genomes (Fang et al., 2012). In bacteria, 6mA was initially reported to be part of restriction-modification (R-M) systems—bacterial defense mechanisms against phages and plasmids that are able to distinguish between host and invader DNA (Arber and Dussoix, 1962). Specifically, the presence of 6mA in the host prevents the digestion of its genome by DNA methylation-sensitive restriction enzymes. In contrast, foreign unmethylated DNA lacks the protection and is readily degraded when entering the cells. R-M system-positive strains are equipped with DNA methyl-transferase and endonuclease counterparts with common sequence recognition motifs.

However, the fact that other methyl-transferases lack a restriction enzyme counterpart and that m6A is important for viability in specific bacterial strains suggests a defense-independent function. Specifically, adenine methylation is established as a bacterial epigenetic mark. Exemplary, solitary adenine methylases, such as Dam in *E. coli*, are involved in DNA replication, wherein sister-strand synthesis can only be

initiated in the presence of methylated adenine at replication origin (Wion and Casadesús, 2006). Dam-mediated methylation also regulates replication initiator factors.

6mA guides the discrimination between original and newly synthesized DNA strand after replication. As de novo adenine methylation is delayed during the cell cycle, the newly synthesized strand is recognized by repair enzymes and the Dam motif enables endonuclease processing with subsequent repair processes (Wion and Casadesús, 2006). Adenine methylation has further functional implication in the cell cycle, repression of transposable elements, and gene regulatory processes (Fang et al., 2012). 6mA also reduces the stability of base pairings, hence favoring transcriptional initiation by lowering the energy to open DNA duplexes. Dam activity can be hindered by binding of competing proteins, resulting in the formation of non-methylated sites. Strikingly, the protection from methylation is an inherited state that, however, can be modified by environmental conditions (Wion and Casadesús, 2006). Thus, adenine methylation displays similar characteristics in prokaryotes as cytosine methylation does in eukaryotes, further underscoring its importance throughout generations.

Adenine Methylation: An Evolutionary Conserved Mechanism

Although some studies hypothesized the presence of 6mA in eukaryotic genomes decades ago, its implication in epigenetics in eukaryotes remains elusive (Ratel et al., 2006). Compared to the highly abundant 5mC in the eukaryotic kingdom, levels of 6mA were suggested to be minimal and thus only detectable by highly sensitive technologies. Nevertheless, several studies reported the presence of 6mA in eukaryotic genomes, particularly in ciliates, chlorophyte algae, and dinoflagellates (Achwal et al., 1983; Gommers-Ampt and Borst, 1995; Ratel et al., 2006). In certain cases, 6mA exists in substantial amounts, with 0.5%–10% of adenines being methylated.

Sequence analysis predicted the presence of adenine methyl-transferases and demethylases in several eukaryotic organisms (Iyer et al., 2011) (Figure 1A). The presence of methyl-transferase

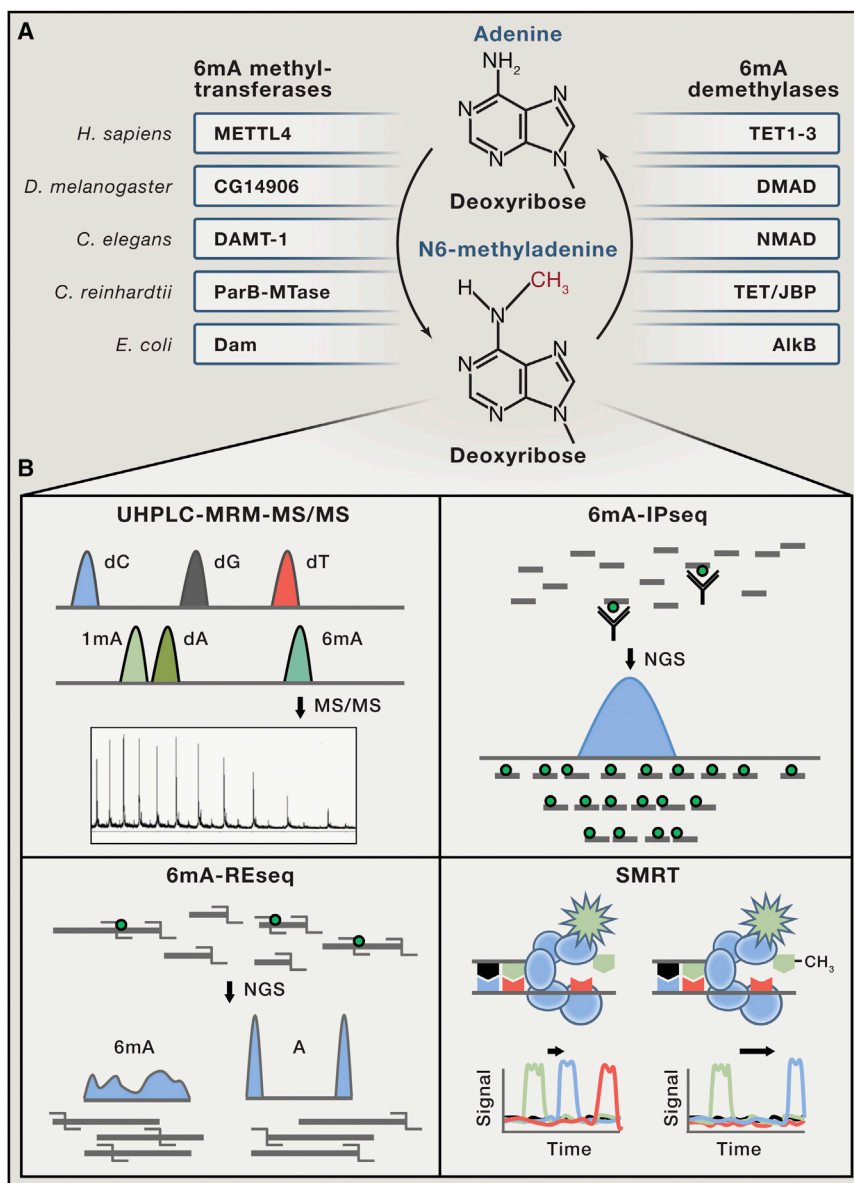


Figure 1. Processing and Detection of N6-Methyladenine

(A) Adenine bases of DNA are modified by N6-methyladenine (6mA) methyl-transferases and 6mA demethylases. The modifying enzymes are conserved in all super-kingdoms of life, with putative activity also in *Homo sapiens* (TET1-3 proteins have so far proven activities as 5mC oxidases).

(B) Methyladenine is detectable by chromatography-based technologies, such as the ultra-high performance liquid chromatography-triple quadrupole mass spectrometry coupled with multiple-reaction monitoring (UHPLC-MRM-MS/MS) method or sequencing approaches. For the specific quantification of methyladenine, next-generation sequencing (NGS)-based strategies are coupled with immunoprecipitation of 6mA (6mA-IPseq) or restriction enzyme guidance (6mA-REseq). Direct quantification at base-pair resolution is enabled by third-generation sequencing methods, such as the single-molecule real-time (SMRT) technology, wherein variant enzyme kinetics identify modified DNA bases.

had yet to be identified (Hattman et al., 1978). Using sequencing-based mapping strategies, Fu et al. produce the first genome-wide reference map for methyladenine in *C. reinhardtii* (Fu et al., 2015). Moreover, the authors provide evidence for an epigenetic function in transcriptional regulation. After confirming abundant 6mA levels by highly sensitive liquid-chromatography and mass-spectrometry methodologies (Figure 1B), they show that 6mA levels are stable and inherited during multiple replication phases. Immunoprecipitation-based sequencing strategies (Figure 1B) identify sequence motifs susceptible to undergo adenine methylation, which are different from the prokaryotic consensus sequences. Subsequently, restriction enzyme-guided resequencing produces a 6mA reference methylome of *C. reinhardtii* at base-pair

orthologs within transposable elements led to the hypothesis of a *cis*-acting control mechanism to secure host genome integrity. Consistently, such a mechanism was identified in *E. coli*, suggesting a conserved function of 6mA as safeguard of the genome (Roberts et al., 1985).

Now, three studies in this issue of *Cell* report the presence of 6mA in three different eukaryotic genomes—*Chlamydomonas reinhardtii*, *Caenorhabditis elegans*, and *Drosophila melanogaster*—with putative epigenetic function (Zhang et al., 2015; Greer et al., 2015; Fu et al., 2015). The authors present evidence for spatiotemporal-regulated 6mA modifications during development. Moreover, 6mA is associated to gene regulatory events.

The green alga *C. reinhardtii* has long been reported to harbor substantial levels of 6mA, but its spatial distribution and function

had yet to be identified (Hattman et al., 1978). Intriguingly, although the methyl-transferase consensus sequence is equally distributed in the genome, 6mA is highly enriched at gene promoters but depleted at the transcription start sites. Consistently, 6mA profiles reveal periodic patterns of 130–140 bp distances and hence a potential association to nucleosome positioning at promoter regions. The presence of 6mA at gene promoters is positively correlated with increased transcriptional activity.

While adenine methylation has been previously described in *C. reinhardtii*, its presence in 6mA *C. elegans* has not been reported despite the presence of putatively active methyl-transferases in the worm genome. Greer et al. now report 6mA to be present in *C. elegans* and functionally involved in epigenetic transgenerational inheritance (Greer et al., 2015). In *C. elegans*, mutants lacking histone demethylase *spr-5*, responsible for

dimethylation of the histone H3 at lysine 4, represent a paradigm of inheritance. Although no phenotype is detectable in early generations, the mutant worms become progressively infertile in later generations, accompanied by increasing histone H3 methylation levels. Surprisingly, Greer et al. now describe that *spr-5* mutants reveal elevated levels of 6mA, accumulating during generations. 6mA in *C. elegans* is shown to be added by the newly identified DNA N6-adenine methyl-transferase 1 (DAMT-1) and dynamically removed by the N6-methyladenine demethylase 1 (NMAD-1). Strikingly, mutations in NMAD-1 lead to accelerated accumulation of 6mA and, moreover, speed up the sterility phenotype in *nmad-1* and *spr-5* double-knockout worms.

Overall, 6mA in *C. elegans* is rather low in wild-type animals (0.025%) but is increased 10-fold in *spr-5* mutant animals. It is noteworthy that, unlike in flies (see below), adenine methylation in *C. elegans* is ubiquitously present in all cell types. Technically, 6mA is determined by different technologies, ranging from global to base-pair resolution profiles using single-molecule real-time (SMRT) sequencing (Figure 1B). Particularly, the latter approach leads to the identification of specific sequence motifs, suggesting a locally regulated deposition of 6mA. However, its functional role remains elusive. Future functional genomics approaches, including a systematic integration of transcriptional profiles, are needed.

The absence of conclusive evidence for cytosine or adenine methylation in *D. melanogaster* has led to the hypothesis that gene regulation takes place without DNA modifications. However, as 6mA is present in eukaryotes at very low levels, Zhang et al. speculate that an impaired function of the putative DNA demethylase DMAD (DNA 6mA demethylase) leads to detectable 6mA in *D. melanogaster* (Zhang et al., 2015). Indeed, using highly sensitive methods (Figure 1B), the authors identify adenine methylation, predominantly in very early developmental stages of the fly embryos (0.07%), but also in somatic cell types. The late-embryo extracts also exhibit elevated demethylating activity compared with early stages.

Demethylation dynamics could be associated with the TET-like protein DMAD, which is dynamically regulated during development. Moreover, DMAD modifies 6mA levels *in vitro* and *in vivo*, and altered demethylase activity leads to increased embryo lethality. 6mA is also detectable in somatic tissue, particularly in ovary and brain cells. Here, 6mA is restricted to certain cell types, being highly abundant in germline cells while losing intensities during germ cell differentiation. In line with these results, DMAD levels increase during egg differentiation, and DMAD mutants present elevated 6mA levels in their ovaries, accompanied by a higher number of undifferentiated cells. Furthermore, high levels of DMAD in brain suggest an antagonistic function in methyl-transferase activities and a dominant suppression of 6mA levels in neurons. 6mA is determined to be enriched in transposon gene bodies, with a putative function in transcriptional activation during early embryonic stages and in undifferentiated cell types.

From the current 6mA knowledge, *C. elegans* and *D. melanogaster* do not present methylcytosine in their genomes. Although the existence of 5mC in *Drosophila* was under controversial discussion for years, recent studies using whole-genome bisulfite sequencing mostly excluded the presence of

5mC in *D. melanogaster* DNA sequence (Raddatz et al., 2013). Hence, the studies by Greer et al. and Zhang et al. suggest 6mA as the unique DNA methylation modification and potentially functional epigenetic mark in *C. elegans* and *D. melanogaster*, respectively. Although the global levels of 6mA are rather low, its local enrichment and sequence specificity point to regulated processing throughout development and differentiation. Future studies need to further establish its role as epigenetic mark and its function in gene regulation.

However, 6mA and 5mC have been described to co-exist in the *C. reinhardtii* genome. Consistently, methyl-transferases and demethylases are conserved in the green alga (Iyer et al., 2011) (Figure 1A). Now, base-pair resolution landscapes of both DNA modifications in *C. reinhardtii* reveal a likely complementary function of 6mA and 5mC, indicated by their spatial separation in the genome (Fu et al., 2015). While 5mC is enriched at the gene bodies of lowly expressed transcripts, 6mA accumulates at the promoter region of highly active genes. It is remarkable that 5mC in *Chlamydomonas* exists at lower levels than observed in higher eukaryotes and is not restricted to CpG motifs (Feng et al., 2010). Taken together, the evidence suggests 6mA to represent an active epigenetic mark in *C. reinhardtii*, while 5mC is likely to be involved in processes downstream of transcriptional initiation.

Intriguingly, although cytosine methylation represents by far the dominant DNA modification in *Homo sapiens*, the machinery to modify adenine nucleotides is conserved during evolution. In this regard, the methyl-transferase-like 4 (METTL4) is similar to DAMT-1 in *C. elegans* (Greer et al., 2015) (Figure 1A). Moreover, active demethylases of the TET family proteins, such as DMAD in *D. melanogaster*, exhibit specificity for methyl adenines and thus might also be implicated in 6mA dynamics in higher eukaryotes (Iyer et al., 2011). In this regard, early studies also reported 6mA in human tissue, specifically placenta (Achwal et al., 1983). However, the presence and function of the adenine code in mammals need to be confirmed by applying novel ultra-sensitive detection technologies (Figure 1B). These technologies will play a key role in improving our understanding on the complexity of DNA modifications in the biology of eukaryotic life and will be discussed below.

The detection of 6mA in human placenta encourages speculations of a specialized function of adenine methylation in specific cell types. Taking into account the mutagenic nature of 5mC, continuously dividing cell types, such as adult stem cells, might have conserved an epigenetic mechanism that better supports the integrity of the DNA template. 6mA presents a potential alternative to 5mC to avoid the accumulation of *de novo* mutations in the immortal DNA strand. In line, 6mA is determined to be highly abundant in early stages of development and undifferentiated reproductive tissue in *D. melanogaster*, supporting the hypothesis of an epigenetic mark with restricted function in pluripotent cell types (Zhang et al., 2015).

Sensitive Detection of Adenine Methylation of DNA

Many of the questions that we have now for 6mA remained open not so long ago for 5mC and 5hmC. For these two cases, the development of bisulfite sequencing and other genome-scale analyses has provided many of the requested answers. Though

the same user-friendly powerful technologies does not exist for 6mA, there are already promising tools to entangle the presence and role of this enigmatic modification in eukaryotes (Figure 1B). Let's briefly summarize them.

Ultra-High Performance Liquid Chromatography-Triple Quadrupole Mass Spectrometry

This approach allows the sensitive detection of nucleotide modifications, such as 5mC and 6mA, at very low abundance (Ito et al., 2011). Briefly, the digested DNA is separated by reverse-phase ultra-high performance liquid chromatography (UHPLC) coupled with mass spectrometry detection using tandem mass spectrometers (MS/MS). Following detection of specific nucleotide modifications, quantification is achieved using a standard curve that is simultaneously analyzed in the sample of interest. It is important to discard any potential contamination from Mycoplasma or bacterial DNA.

6mA-Immunoprecipitation Sequencing

Immunoprecipitation coupled with next-generation sequencing was previously established for 5mC detection in mammalian genomes (Weber et al., 2005). 6mA-immunoprecipitation sequencing (6mA-IPseq) utilizes a specific antibody for methyladenine to enrich modified fragments from the sequencing library. Following the alignment of sequencing reads to the reference genome, 6mA-modified regions present enriched mapping frequencies. 6mA-IPseq allows charting the spatial distribution of the epigenetic mark. Subsequent sequence enrichment analysis can also point to consensus recognition motifs for the adenine methyl-transferases.

Restriction Enzyme-Based 6mA Sequencing

Restriction enzyme-based 6mA sequencing (6mA-REseq) relies on the determination of consensus target sequences of adenine methylation, followed by the identification of restriction enzymes with respective recognition sites and sensitivity for the DNA modification (Fu et al., 2015). Technically, genomic DNA is fragmented with a 6mA-sensitive enzyme, followed by random shearing of the template. It results in an enrichment of unmethylated (digested) sequence motifs at the ends of the sequencing reads. Conversely, methylated adenine prevents digestion and is enriched in inner fractions of the reads. Consequently, 6mA levels are readily inferred from the relative position of the restriction enzyme consensus sequence.

Single-Molecule Real-Time Sequencing

Initial genome-wide methyladenine maps at base-pair resolution were obtained in *E. coli* genomes using single-molecule real-time (SMRT) sequencing (Clark et al., 2012; Fang et al., 2012; Murray et al., 2012). SMRT, a third-generation sequencing technique, is based on the processing of fluorescence-labeled nucleotides by DNA polymerases. The fluorescence label is not incorporated in the de novo synthesized strand but is cleaved away during the process. Meanwhile, the label emits light that is captured in the nanophotonic visualization chamber. High-fidelity polymerases are capable of synthesizing long continuous strands at a high speed, allowing a fast sequencing process and high read lengths. Importantly, the incorporation of a modified nucleotide, such as 6mA, presents different kinetics compared with unmodified adenine, allowing the direct inference of the modification status of each base.

Conclusions

m6A is a covalent modification of DNA that exerts an essential role in bacteria, where it is associated with genome protection via R-M systems. Furthermore, formation of m6A plays roles in bacterial DNA replication, mismatch repair, and gene transcription. Its presence in the genomes of several eukaryotes reinforces the notion that m6A is widespread and suggests its still unknown activities. The accompanying articles in this issue of *Cell* describe a transcriptional regulatory role for m6A in *Chlamydomonas*, and its detection, although at low levels, in *D. melanogaster* and *C. elegans* indicates an expanded function for 6mA. The development of improved technologies to unambiguously quantify and characterize 6mA in different biological contexts will be a necessary step in this exciting journey.

REFERENCES

- Achwal, C.W., Iyer, C.A., and Chandra, H.S. (1983). *FEBS Lett.* 158, 353–358.
- Arber, W., and Dussoix, D. (1962). *J. Mol. Biol.* 5, 18–36.
- Clark, T.A., Murray, I.A., Morgan, R.D., Kislyuk, A.O., Spittle, K.E., Boitano, M., Fomenkov, A., Roberts, R.J., and Korlach, J. (2012). *Nucleic Acids Res.* 40, e29.
- Fang, G., Munera, D., Friedman, D.I., Mandlik, A., Chao, M.C., Banerjee, O., Feng, Z., Losic, B., Mahajan, M.C., Jabado, O.J., et al. (2012). *Nat. Biotechnol.* 30, 1232–1239.
- Feng, S., Cokus, S.J., Zhang, X., Chen, P.-Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H., Halpern, M.E., et al. (2010). *Proc. Natl. Acad. Sci. USA* 107, 8689–8694.
- Fu, Y., Luo, G.-Z., Chen, K., Deng, X., Yu, M., Han, D., Hao, Z., Liu, J., Lu, X., Dore, L.C., et al. (2015). *Cell* 161, this issue, 879–892.
- Gommers-Ampt, J.H., and Borst, P. (1995). *FASEB J. Off.* 9, 1034–1042.
- Greer, E.L., Blanco, M.A., Gu, L., Sendinc, E., Liu, J., Aristizabal-Corrales, D., Hsu, C.-H., Aravind, L., He, C., and Shi, Y. (2015). *Cell* 161, this issue, 868–878.
- Hattman, S., Kenny, C., Berger, L., and Pratt, K. (1978). *J. Bacteriol.* 135, 1156–1157.
- Heyn, H., and Esteller, M. (2012). *Nat. Rev. Genet.* 13, 679–692.
- Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). *Science* 333, 1300–1303.
- Iyer, L.M., Abhiman, S., and Aravind, L. (2011). *Prog. Mol. Biol. Transl. Sci.* 101, 25–104.
- Kohli, R.M., and Zhang, Y. (2013). *Nature* 502, 472–479.
- Murray, I.A., Clark, T.A., Morgan, R.D., Boitano, M., Anton, B.P., Luong, K., Fomenkov, A., Turner, S.W., Korlach, J., and Roberts, R.J. (2012). *Nucleic Acids Res.* 40, 11450–11462.
- Raddatz, G., Guzzardo, P.M., Olova, N., Fantappiè, M.R., Rampp, M., Schaefer, M., Reik, W., Hannon, G.J., and Lyko, F. (2013). *Proc. Natl. Acad. Sci. USA* 110, 8627–8631.
- Ratel, D., Ravanat, J.-L., Berger, F., and Wion, D. (2006). *BioEssays* 28, 309–315.
- Roberts, D., Hoopes, B.C., McClure, W.R., and Kleckner, N. (1985). *Cell* 43, 117–130.
- Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., and Schübeler, D. (2005). *Nat. Genet.* 37, 853–862.
- Wion, D., and Casadesús, J. (2006). *Nat. Rev. Microbiol.* 4, 183–192.
- Zhang, G., Huang, H., Liu, D., Cheng, Y., Liu, X., Zhang, W., Yin, R., Zhang, D., Zhang, P., Liu, J., et al. (2015). *Cell* 161, this issue, 893–906.

Making Sense of Transcription Networks

Trevor R. Sorrells¹ and Alexander D. Johnson^{1,*}

¹Department of Biochemistry & Biophysics and Department of Microbiology & Immunology, Tetrad Graduate Program, University of California, San Francisco, San Francisco, CA 94158, USA

*Correspondence: ajohnson@cgl.ucsf.edu
<http://dx.doi.org/10.1016/j.cell.2015.04.014>

When transcription regulatory networks are compared among distantly related eukaryotes, a number of striking similarities are observed: a larger-than-expected number of genes, extensive overlapping connections, and an apparently high degree of functional redundancy. It is often assumed that the complexity of these networks represents optimized solutions, precisely sculpted by natural selection; their common features are often asserted to be adaptive. Here, we discuss support for an alternative hypothesis: the common structural features of transcription networks arise from evolutionary trajectories of “least resistance”—that is, the relative ease with which certain types of network structures are formed during their evolution.

Introduction

The complexity of cells continues to fascinate scientists. Two broad views are often advanced to account for such complexity. In one, it is assumed that any complexity must necessarily benefit the cell. Some cell and molecular biologists go even further and discuss how a particular mechanism was “designed” by evolution to be perfectly matched to its task. As with a machine, it is assumed that every molecular nut and bolt must have a purpose. Because this view seems intuitive and relatively simple (after all, examples abound of animals, plants, and microbes adapted to their environments), it is often invoked to explain any aspect of cell and molecular biology. A different view, the one we elaborate here, is embodied in Dobzhansky’s famous line, now a cliché, “nothing in biology makes sense except in the light of evolution.” According to this view, any rationalization of a modern cellular mechanism depends critically on understanding its evolutionary history. We argue that this emphasis on evolutionary history is especially appropriate for analyzing transcription circuits and for rationalizing their structures. This view has explanatory power in that it can readily account for some of the more bewildering and counterintuitive features of modern transcription circuits; it also gives us insight into the best ways to describe and study such circuits.

In this Perspective, we first review common features of transcription network structures—observed across diverse species—and argue that these similarities cannot be the result of descent from a single ancestral circuit possessing these characteristics. Next, we consider key biochemical and biophysical properties of transcription regulators and *cis*-regulatory sequences that make certain evolutionary pathways much more probable than others, in part because they circumvent fitness barriers. Finally, we argue that many aspects of transcription circuits, particularly those that seem overly complex and counterintuitive, can be understood as relatively crude products of high-probability evolutionary trajectories rather than as highly optimized, specific solutions.

The arguments discussed in this perspective rely heavily on prior ideas advanced by evolutionary biologists, particularly

those ideas concerning the role of non-adaptive mutations in generating complexity (Covello and Gray, 1993; Doolittle, 2013; Force et al., 1999; Gray et al., 2010; Lukeš et al., 2011; Lynch, 2007a, 2007b, 2014; Stoltzfus, 1999; Zuckerkandl, 1997). Although sometimes dismissed as unimportant (or uninteresting), non-adaptive mutations can have a profound role in generating evolutionary novelty. Of particular importance is the idea, sometimes called “constructive neutral evolution,” that changes that arise neutrally can open up new evolutionary pathways; in some cases, changes that arose non-adaptively can become essential for function if they are incorporated into subsequent layers of evolutionary change. Through this sequence of events, molecular and organismal complexity can be increased through non-adaptive mutations. As we discuss, the biochemical and biophysical properties of transcription network components support the idea that their evolutionary trajectories—which depend on mutation, selection, and genetic drift—lead to specific types of structures. Because their components are highly conserved across eukaryotes, we argue that it is inevitable that networks across a wide variety of species tend to converge on similar structures. We propose that these common structures are not likely to represent optimized solutions but are, in a sense, “default” evolutionary products.

Depictions of Transcription Networks

For the most part, genome-wide studies of transcriptional network structures have been largely descriptive, often culminating in large “hairball” diagrams such as those depicted in Figure 1. Their complexity has made it difficult to formulate simple conclusions regarding the logic or outputs of these networks, particularly since quantitative parameters and dynamic measurements are typically lacking.

Although there are many components of gene expression networks, we will focus here on only two key elements, transcription regulators and *cis*-regulatory sequences. We define transcription regulators as sequence-specific DNA-binding proteins that control the transcription of specific genes by binding to *cis*-regulatory sequences, short (typically 6–15 nucleotides) DNA

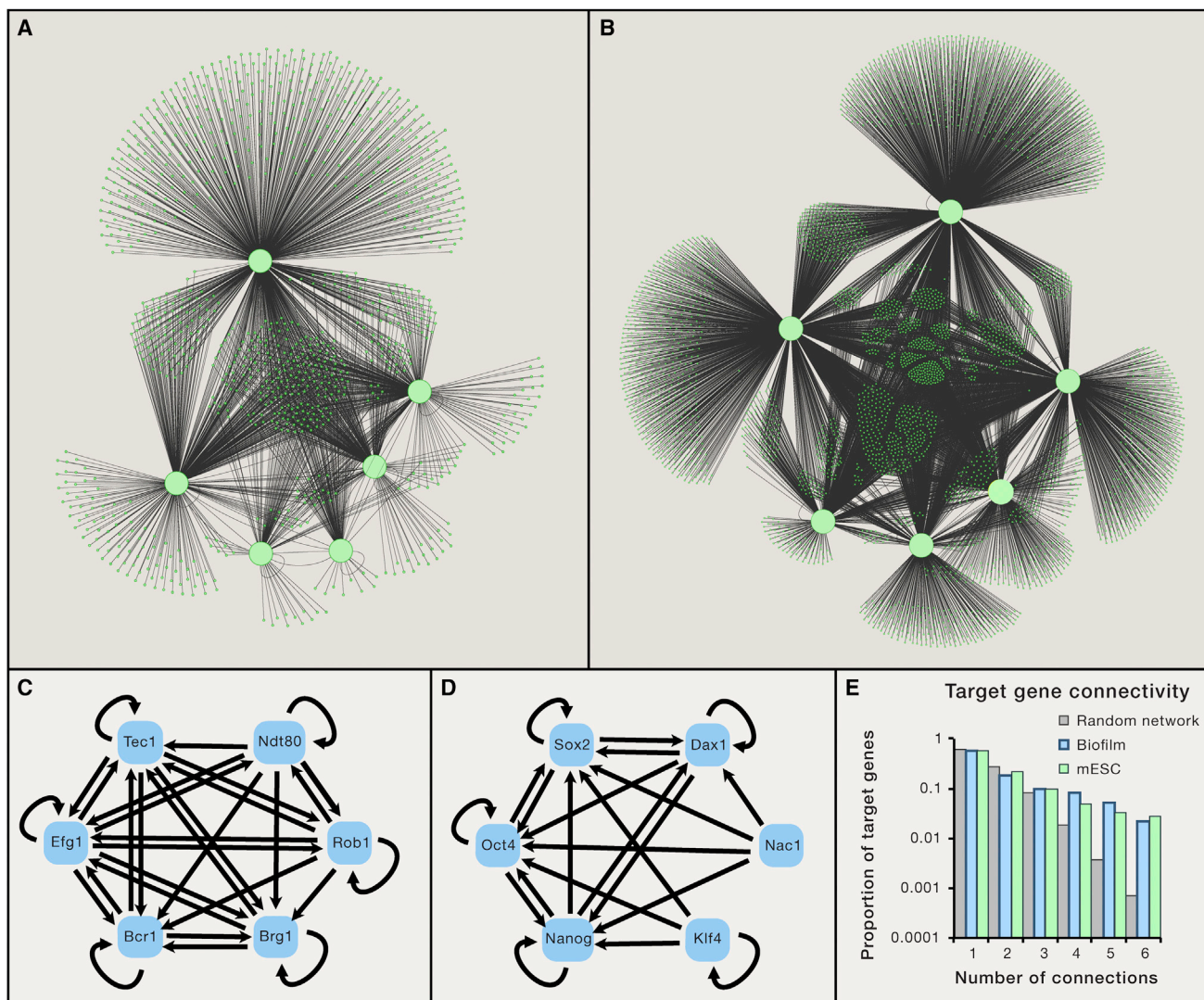


Figure 1. Typical Depictions of Transcription Regulatory Networks

(A and B) (A) The *C. albicans* biofilm network (Nobile et al., 2012) and (B) the *M. musculus* embryonic stem cell network (Kim et al., 2008) are depicted as graphs where balls represent genes and lines represent the binding of transcription regulators to intergenic regions. Master transcription regulators (defined in the text) are shown as large balls, and “target genes” are shown as small balls. For the stem cell network, only the six most heavily connected transcription regulators are shown.

(C and D) Close-up of the core of each network, showing only the binding connections between the master transcription regulators. Directionality of the connection is indicated by arrows. Note that the arrows refer only to binding connections and do not imply that the connection activates the recipient gene. (C) *C. albicans* biofilm, (D) mouse stem cell networks.

(E) The degree of connectivity for nodes in the two networks. The two biological networks show a larger proportion of nodes with high connectivity than would be found in a random network (Lee et al., 2002).

sequences. It is the distribution of these *cis*-regulatory sequences across the genome that largely specifies the time, place, and rate of each gene’s transcription; this information is “read” by transcription regulators, whose binding to DNA specifies, often through a complex series of downstream steps, the rate of transcription of the gene. Although in many eukaryotic species, *cis*-regulatory sequences are typically located within several thousand nucleotide pairs of the genes they control, in plants and animals, they can be spread out over hundreds of thousands of nucleotide pairs. Nearly all eukaryotic genes are directly controlled by more than one transcription regulator,

and most genes respond to dozens of regulators, specified by the identity and arrangement of their *cis*-regulatory sequences. We also know, from decades of “promoter bashing” experiments, that *cis*-regulatory sequences can be moved from one gene to another (and from one species to another) and still retain much of their specificity to direct transcription. Finally, transcription regulators typically bind cooperatively to DNA, a fundamental property that, as we shall discuss, has important implications for network evolution.

Many additional proteins besides transcription regulators are needed to transcribe a gene (for example, RNA polymerase

Table 1. Metrics Comparing *C. albicans* Biofilm and Mouse Embryonic Stem Cell Networks

	Biofilm	mESC
Master transcription regulators	6	6
Connections	2018	7234
Target genes	1037	3968
Fraction of genome in network	0.17	0.21
Binding feed-forward loops	3145	6886
Non-functional binding events	1207	unknown

Connections were determined by whole-genome chromatin immunoprecipitation and the table values are simply total counts without regard for the number of significant figures. As such, they should be regarded as approximate, particularly since there is no evidence that either circuit is completely described. “Non-functional binding” was defined as genes whose expression does not change when a direct regulator is deleted from the genome.

and chromatin remodeling complexes), but the specification of which genes are transcribed in which tissues is determined by direct binding connections between transcription regulators and genes (or more precisely, the *cis*-regulatory sequences of that gene). This information is summarized in diagrams such as those in Figure 1.

If a given transcription regulator occupies the *cis*-regulatory sequences associated with a gene in vivo (as determined, for example, by a chromatin immunoprecipitation experiment), we will refer to that gene as a target gene of the transcription regulator. We realize that this convention does not require that the binding of the regulator to DNA be proven to be functional in the organism. There are three reasons for nonetheless including these connections in diagrams such as those in Figure 1. (1) The “function” of a given connection has been demonstrated in only a small number of cases; for the great majority of reliable binding data, no direct test has been performed. (2) Although many approaches (e.g., conservation across species or experimental mutation of the *cis*-regulatory sequence) can provide strong evidence for function, it is impossible to rigorously establish that a binding connection is non-functional under all possible conditions. (3) The DNA-binding properties of transcription regulators predict that, in vivo, there will be some degree of non-functional binding (Lin and Riggs, 1975). Such “non-functional” binding events are nonetheless real properties of evolving transcription networks.

Depictions of transcription networks based on these conventions often show “master transcription regulators” and target genes as nodes (balls) and regulatory interactions as edges (lines) between these nodes (Figures 1A and 1B). Although the term master transcription regulator is used in many different ways in the literature (Chan and Kyba, 2013), we define it, for the purpose of this article, as a transcription regulator (1) whose presence is required to carry out the specific biological process controlled by the network and (2) whose ectopic expression alone or in combination with other regulators can trigger the biological process even in the absence of the ordinary developmental or environmental signals (Halder et al., 1995; Takahashi and Yamanaka, 2006; Tapscott et al., 1988; Tursun et al., 2011; Vierbuchen et al., 2010; Zordan et al., 2007).

Common Features of Transcription Networks

We first compare two transcription networks from two different species that coordinate two different biological processes, but were deduced by similar methodologies. The network specifying the embryonic stem cell state (pluripotency) was chosen because it has been studied extensively by numerous labs and is supported by multiple studies (Boyer et al., 2005; Kim et al., 2008). For comparison, we chose the circuit controlling biofilm development in the pathogenic yeast *C. albicans*, a network this lab has studied extensively (Nobile et al., 2012). The two networks are depicted in Figures 1A and 1B using a similar graphical format.

These two circuits were chosen, in part, because they might be expected from first principles to have little in common. Mammals and yeast diverged from a common ancestor ~1.5 billion years ago (Wang et al., 1999), and there is little conceptual similarity between biofilm formation and pluripotency. Moreover, the two networks appear to have evolved independently, well after the two lineages split (see below). Yet, the overall structures of the two networks, as depicted in the figure, appear remarkably similar. Both *C. albicans* biofilm development and mouse embryonic stem cell pluripotency are controlled by a set of master transcription regulators that form binding connections among themselves (Figures 1C and 1D) and to the regulatory regions of over a thousand target genes, with multiple master regulators typically binding to the same targets (Figures 1A, 1B, and 1E and Table 1). In both cases, a substantial proportion of the target genes are other transcription regulators, indicating substantial indirect regulation of additional genes. The *C. albicans* genome is significantly smaller than the mouse genome, yet each network comprises about one-fifth of the coding genes in their respective genomes.

Although the two networks control very different processes, their master regulators have similar properties. In both networks, these regulators contain ordinary sequence-specific DNA-binding domains such as homeodomains, MADS domains, and zinc fingers (Weirauch and Hughes, 2011). In some cases, the *cis*-regulatory sequence recognized by a given transcription regulator has not changed significantly since the divergence of yeast and mammals (Hayes et al., 1988). Moreover, transcription regulators from one species (e.g., Gal4 from brewer's yeast) can control transcription in many different species (e.g., Fischer et al., 1988; Kakidani and Ptashne, 1988). A master regulator for a given process is therefore distinguished by its connections, not its structure.

Key aspects of the *C. albicans* biofilm circuit were formed well after *C. albicans* diverged from closely related, non-pathogenic yeasts (Nobile et al., 2012), providing additional support for the conclusion that the structure of the yeast and mouse networks evolved independently—even though the transcription regulators that would become master regulators for these two networks were present in the common ancestor of both species. These ideas are consistent with the generalization that, although the transcription regulators and their recognition sequences are often deeply conserved, transcriptional networks themselves are rewired at a rapid pace during evolution (reviewed in Li and Johnson, 2010; Tuch et al., 2008b; Weirauch and Hughes, 2010; Wray et al., 2003). Like most generalizations in biology, this one has important exceptions. See, for example, Baker et al. (2011) and

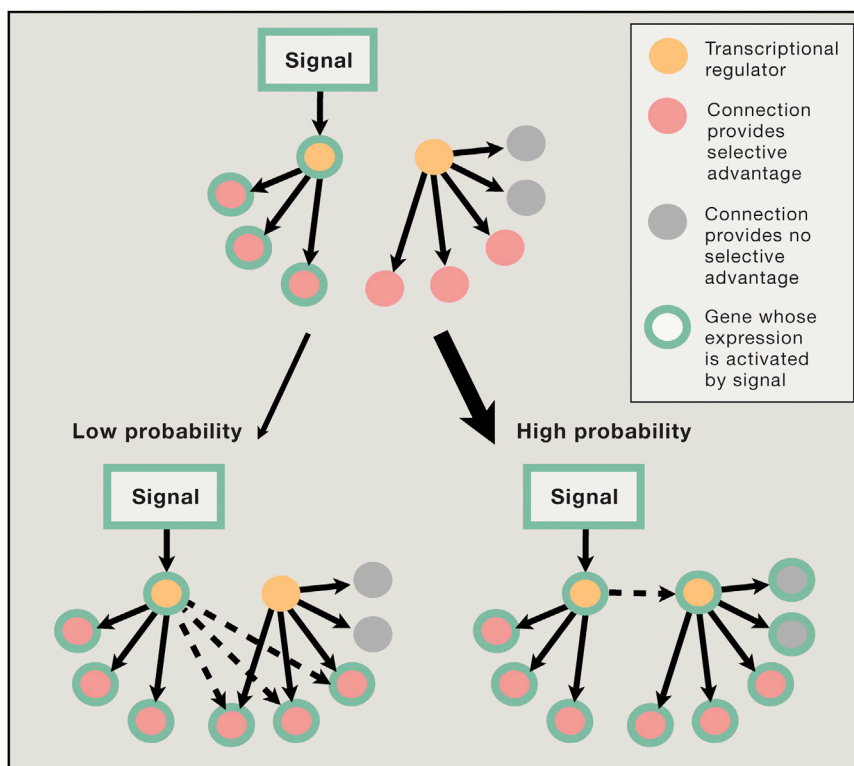


Figure 2. Pathways for Evolving a New Transcriptional Response to a Signal

In this hypothetical scenario, incorporation of three additional genes into the signaling pathway confers a selective advantage. Two alternative paths are possible: (1) the genes could be incorporated one by one through independent changes in their *cis*-regulatory sequences (gene-by-gene model). (2) The new genes could be incorporated through a single incorporation of the transcription regulator that already controls them (regulator-first model). If the incorporation of multiple target genes is needed to confer an increase in fitness, gain of regulation of the transcription regulator will be more probable than the gain of each individual target. As the number of target genes increases, the difference in probability will be greater. Note that the second scenario will likely incorporate additional genes non-adaptively.

virtually all other organisms), and we suggest that the available trajectories for evolutionary change are based on these simple properties coupled with the avoidance of fitness barriers. According to this view, the similarities among independently derived transcription networks arise primarily from the “low-energy” pathways of evolution—rather than the selective pressures specific to one circuit

Sayou et al. (2014) for cases in which the DNA-binding specificity of a regulator has changed dramatically over relatively short periods of evolutionary history. In any case, it is highly unlikely that any of the connections between regulators and target genes in the mouse pluripotency and yeast biofilm networks are conserved from a common ancestor, despite the deep conservation of the DNA-binding properties of the master transcription regulators.

If transcription networks evolve rapidly, why do the embryonic stem cell and biofilm networks appear structurally similar? One hypothesis is that elaborate and interconnected networks such as these represent optimized solutions for organizing biological processes. According to this view, the similarities between these networks result primarily from selection and reflect the same underlying requirements for transcriptional logic—for example, modularity or robustness. Some features of the circuits (for example, the large number of direct and indirect feedback loops) may well reflect these requirements in a general way, but the similarities seem too great to be readily explained this way. We propose instead that circuit architecture is dominated by severe constraints on the evolutionary trajectories available for network evolution. Allowable trajectories, we argue, must (1) be probable from a biochemical and biophysical standpoint and (2) avoid fitness barriers; that is, the allowable trajectories will typically not pass through stages in which the circuit becomes broken and non-functional (Carroll, 2008; Stern and Orgogozo, 2009; Wagner, 2003).

The components of circuits (DNA-binding proteins and *cis*-regulatory sequences) and their properties (for example, cooperative binding) are common to fungi and mammals (as well as

or another. In the following sections, we examine specific properties of networks in more detail and consider the extent to which this idea can account for them.

Size Accrues

One surprising feature of many transcriptional networks is their large size (Borneman et al., 2006; Hernday et al., 2013; Iyer et al., 2001; Junion et al., 2012; Kim et al., 2008; Liang and Biggin, 1998; MacArthur et al., 2009; Mastick et al., 1995; Nobile et al., 2012; Novershtern et al., 2011). As mentioned above, the yeast biofilm network and the mouse embryonic stem cell network, as depicted in Figure 1, incorporate approximately one-fifth of the protein-coding genes in their respective genomes. Although a few examples have been described in which eukaryotic transcription networks appear small (e.g., the mating type specification circuit [Galgoczy et al., 2004] and the galactose regulatory circuit [Ren et al., 2000], both from *S. cerevisiae*), the majority of networks that have been carefully studied using full-genome methods appear larger and more complex than might have been expected.

Why is the typical network so large? In contrast to a model in which every connection in a network serves a specific function in that network, we propose that many target genes in networks are incorporated non-adaptively during the formation of the network. Figure 2 shows a hypothetical example in which a new response to a signal evolves under selection. If there is an advantage of gaining regulation of multiple target genes in response to the signal, it is much more probable to gain a binding site upstream of a single transcription regulator of those genes than to gain binding sites for each individual target gene (Gerhart and

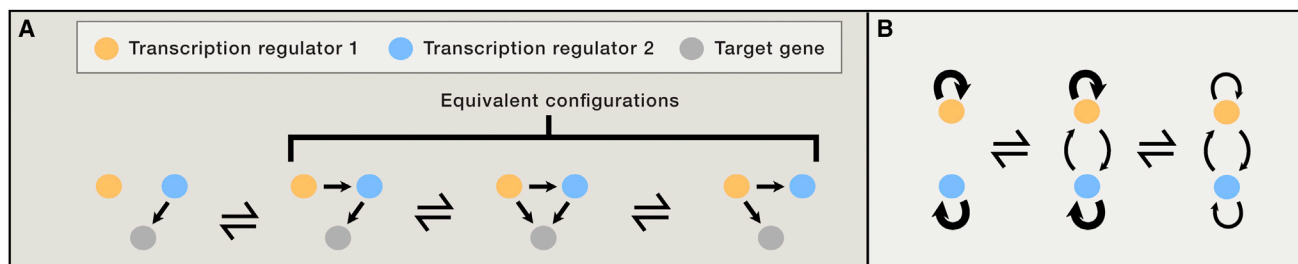


Figure 3. The Tendency for Co-Expressed Regulators to Become Interconnected

(A) Once the orange regulator gains control of the blue regulator, causing them to be expressed at the same time, target genes can, through neutral evolution, rapidly gain and lose binding sites for the two regulators.

(B) Two regulators expressed at the same time each have positive feedback. Subsequently, neutral gains of reciprocal regulation between the regulators can occur while preserving the overall positive feedback control. Over evolutionary time, positive feedback distributed over both regulators (rather than purely autonomous loops) is predicted to occur.

Kirschner, 1997; Raff and Kaufman, 1983). Moreover, because most proteins work in groups, any selective advantage of incorporating a new gene into an existing circuit may not be realized until several genes are brought into the circuit, making the “gene-by-gene” model even less probable. The “regulator-first” model would result in a new regulator being incorporated into the old circuit, along with all the pre-existing target genes of this regulator. Some of these target genes may be extraneous with respect to the new circuit, but, if the original function of the regulator is retained, these connections would nonetheless be maintained by purifying selection. According to this simple idea, newly formed networks would be expected to contain connections nonessential to that network and would therefore be predicted to be larger than strictly necessary.

Experimental evidence suggests that the “regulator-first” scenario is common; that is, networks often form by incorporating new regulators rather than by incorporating individual target genes (Frankel et al., 2012; Monteiro, 2012; Pires et al., 2013). For example, the red wing color in *Heliconius* butterflies takes place through repeated rewiring of the expression pattern of the transcription regulator *optix* rather than one-by-one incorporation of individual target genes (Reed et al., 2011). A second example is found in networks regulating morphological transitions in different yeast species; the regulator Tec1 and its target genes have been incorporated into environmental response networks multiple times (Mösch and Fink, 1997; Nobile et al., 2012; Schweizer et al., 2000). Thus, the regulators-first model of transcription network formation is predicted to lead to the expansion of circuit size beyond that strictly required for the new response. Although this model might be expected to create detrimental pleiotropic effects of expressing many extraneous genes at once, modeling and experimental evidence suggests that this pleiotropy can be alleviated gradually over time (Pavlicev and Wagner, 2012; Qian et al., 2012) or even avoided altogether (Stern and Orgogozo, 2009). It is important to note that these observations, although widely applicable to many eukaryotic species (particularly those with small effective population sizes), do not hold in every case. For example, in some viruses, genomes are compact, regulatory networks are small, and each component and connection contributes to the function of the circuit (Little, 2010).

Gains in Interconnectedness

Another common feature of transcription networks across diverse species is the degree of connectivity between different transcription regulators and between these regulators and their targets (Borneman et al., 2006; Boyle et al., 2014; Junion et al., 2012; Kim et al., 2008; MacArthur et al., 2009; Nobile et al., 2012; Novershtern et al., 2011; Reece-Hoyes et al., 2013). We define this degree as the number of connections made between the master transcription regulators and a given target gene. For example, if a given target gene in the *C. albicans* biofilm network is bound by three different master transcription regulators, the degree of connection of that target gene is three. The degree distributions for the yeast and mouse cases show a similar profile (Figure 1E), one that shows a higher degree of connection than would be predicted for a randomly distributed network (Featherstone and Broadie, 2002; Guelzim et al., 2002).

Rather than speculating what this high degree of interconnectedness might “do for the cell,” we subscribe to the simpler hypothesis that it results from the neutral (i.e., non-adaptive) gains of regulatory connections that inevitably occur over time, particularly in small populations (Lynch, 2007a; Stone and Wray, 2001). This idea can be explained by considering a simple situation, one that would be predicted to arise often by the “regulator-first” model (Figure 3A). Here, one transcription regulator (blue) regulates the target gene (gray). In the regulator-first model, a second transcription regulator (orange) gains control of the blue regulator, and indirectly, the gray target gene. Next, the interconnectedness in this simple scheme would increase if the orange regulator gained direct control of the gray target gene without losing its initial connection.

Why would this happen? It has been argued, using population genetic models, that such connections are predicted to form non-adaptively—that is, without selection for improvement of the circuit (Lynch, 2007a). According to this view, the additional connections do not disrupt the existing regulation, and they arise through random mutations that produce a new DNA-binding site for a transcription regulator. Thus, this increase in total number of connections is predicted to occur, in essence, because nothing stops it. Such a change, even though it arose non-adaptively, can become necessary if subsequent evolutionary changes in the network render its loss detrimental.

Although it might seem counterintuitive that new circuit connections can form non-adaptively, the biochemical features of transcription regulators and *cis*-regulatory sequences predict this. As has been pointed out many times, because *cis*-regulatory sequences are usually short and somewhat degenerate, there is a significant probability that new point mutations will readily create bona fide *cis*-regulatory sequences for existing transcription regulators. Given that target genes often have long intergenic regions in which *cis*-regulatory sequences can function, many target genes would be predicted to develop multiple connections (Lynch, 2007a; Paixão and Azevedo, 2010; Stone and Wray, 2001). Although these additional binding sites may not be under purifying selection (unless the original connection is lost or some other change in the network renders their loss detrimental), they would be predicted to form at a high enough frequency to ensure an appreciable steady-state level of such connections, despite the losses due to mutation.

These same forces are also predicted to lead to the high interconnectedness observed between the master regulators themselves (Figure 1B). Many transcription regulators control their own transcription (Bateman, 1998; Kielbasa and Vingron, 2008; Lee et al., 2002). When two such regulators function at the same time and place (although not necessarily in the same biological function), over time they may acquire regulation of each other through the gain of *cis*-regulatory sequences (Figure 3B). This reciprocal regulation would be redundant (at least in a general sense) with the auto-regulation of each of the transcription regulators themselves and could partially replace it over time, resulting in interlocking, auto-regulatory master regulators of the type we see in Figure 1.

Various types of simulations both support these ideas and highlight additional features that promote high degrees of circuit connectivity. For example, long regulatory regions and high recombination rates promote the evolution of multiple *cis*-regulatory sites by non-adaptive mechanisms (Lynch, 2007a; Ruths and Nakhleh, 2012). Similarly, the greater the permissible degeneracy of *cis*-regulatory sequences, the greater is the probability of multiple connections (Paixão and Azevedo, 2010).

Support for these ideas also comes from direct observation of transcription circuits in various species. First, as previously pointed out, independently evolved circuits show similar, high degrees of connectivity. Recent studies of transcription networks by the Encyclopedia of DNA Elements (ENCODE) project have greatly increased the number of examples where network structure is observed to be highly similar across organisms—in this case, humans, mice, *Caenorhabditis elegans*, and *Arabidopsis thaliana* (Boyle et al., 2014; Stergachis et al., 2014; Sullivan et al., 2014). Although some similarities (e.g., the same master regulator controlling the same biological process in two different species) are clearly conserved from a common ancestor, we argue that similarities in overall network structure are largely due to the pathways we have outlined, in which non-adaptive evolution is a major force.

Second, there are several documented examples in which evolutionary rewiring of an entire network has occurred without apparent changes in the output (Baker et al., 2012; Lavoie et al., 2010; Ludwig et al., 2000; Moses et al., 2006; Schmidt et al., 2010; Tanay et al., 2005; Tsong et al., 2003; 2006). These

studies indicate that, even as the output of a circuit is maintained by stabilizing selection, the individual connections may be free to drift to new configurations.

Third, many connections in networks appear unimportant as assessed by conventional experiments (Fisher et al., 2012; Whitfield et al., 2012). Although it is virtually impossible to prove that a connection is non-functional under all conceivable conditions, several types of experiments suggest that parts of circuits may be functionally unimportant. For example, many direct target genes show no change in transcript levels when a regulator that binds to the gene is deleted or reduced in expression. This behavior describes the majority of the *C. albicans* biofilm network: 60% of binding events do not elicit expression changes when the regulator is deleted, with the provision that biofilms were monitored under a narrow range of conditions. Moreover, many target genes, when deleted, do not appear to compromise the output of the circuit. Although these results can be explained away by circuit compensation, redundancy, inability to monitor a wide variety of conditions, and the like, we suggest it is highly plausible, based on the arguments made above, that many circuit connections simply do not contribute to the output. In any case, many observations made on modern circuits are consistent with a model whereby much of the interconnectedness of transcription circuits has arisen non-adaptively, simply as a consequence of the ease of forming new connections.

Cooperative Binding Produces Connectivity

Cooperative binding is a near-universal feature of eukaryotic transcription regulators, and next we discuss how this property increases the ease of forming new circuit connections and thereby shapes circuit structures. We use the term cooperative binding to mean that the binding of one transcription regulator to a *cis*-regulatory sequence increases the probability that another will occupy a nearby sequence. Mechanistically, this can occur through three distinct means: (1) competitive displacement of nucleosomes, through which binding of one transcription regulator to DNA can increase the accessibility of DNA to a second regulator, thereby increasing its occupancy (Polach and Widom, 1996); (2) a direct, weak, favorable, physical interaction between the two regulators (Johnson et al., 1979); and (3) physical interactions with additional non-DNA-binding proteins that stabilize binding of both of the transcription regulators on DNA (Ptashne and Gann, 2002).

All three forms of cooperative binding would favor the drift of circuits into states of high connectivity by relaxing the *cis*-regulatory sequence requirements needed for a second transcription regulator to be added to a target gene. This idea has an additional implication: cooperativity means that a single change in a *cis*-regulatory sequence or a regulatory protein can establish or eliminate numerous connections. For example, gain of a *cis*-regulatory site for one regulator may allow other regulators to occupy nearby, previously existing weak sites (Figure 4A). Acquisition of a new, favorable protein-protein interaction between transcriptional regulators can have an even more profound effect. Here, cooperative binding can, at least in principle, catalyze the rewiring of an entire set of genes rather than a single gene (Tuch et al., 2008a). In this scenario, the gain of a protein-protein

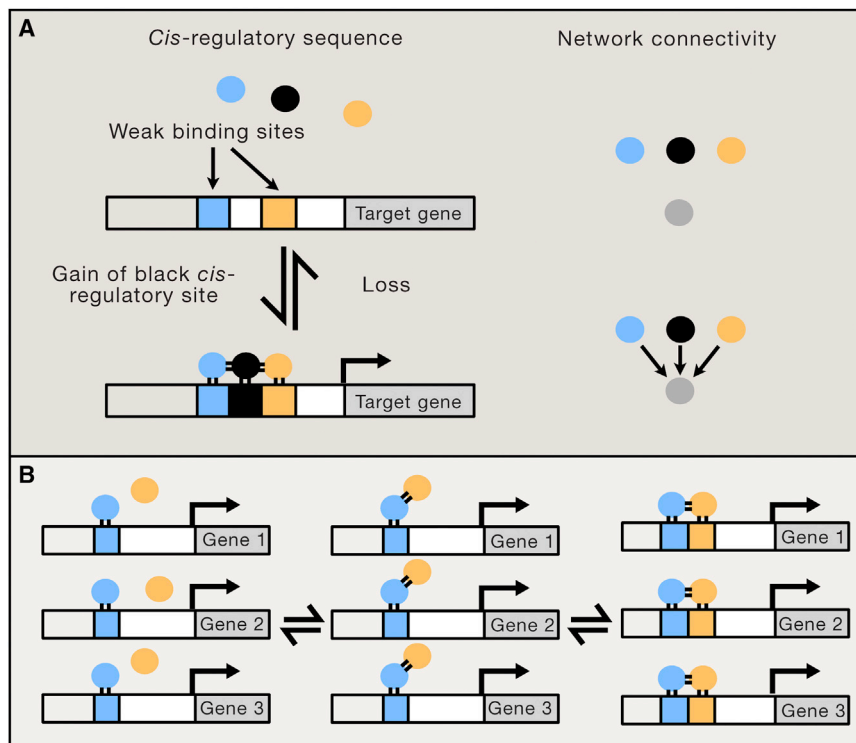


Figure 4. Gain of Multiple Regulatory Connections through Cooperative Binding

(A) Cooperativity between regulators allows binding energy to be shared between protein-DNA and protein-protein interactions. When a strong binding site is gained for one regulator, this may increase the occupancy of regulators on nearby weak binding sites that would otherwise be unoccupied. The effect is a concerted increase in connectivity of that target gene.

(B) The gain of a protein-protein interaction between the blue and orange regulators results in a concerted rewiring of the entire set of genes. As shown in the third panel, direct binding sites for the orange regulator can be gained stepwise at each gene individually without disrupting the circuit. Finally (not shown), the circuit can diversify by moving between equivalent configurations. For example, the ancestral (blue) or derived (orange) connections could be lost without destroying the regulation.

interaction leads to cooperative binding of two regulators when a binding site for only one of the regulators is present (Figure 4B). Following this gain, there can be gene-by-gene gains of *cis*-regulatory sequences for the second regulator. The dually regulated set of genes can then diversify, loosening their connections with the original regulator and strengthening the new ones. In this way, gene sets can be “handed off” from one regulator to another in the course of evolution, a type of change that seems common (Baker et al., 2012; Lavoie et al., 2010; Martchenko et al., 2007; Tanay et al., 2005; Tsong et al., 2006).

The important point is, to influence transcription, transcription regulators must occupy *cis*-regulatory sequence in the cell, but the energy needed for this occupancy can be shared out between protein-DNA and protein-protein interactions. As individual interactions are strengthened and weakened over evolutionary time, the circuit configuration can drift between different “energy-sharing” solutions. Cooperative binding, combined with the ease of strengthening and weakening *cis*-regulatory sequences by random mutations, predicts that networks will drift to high degrees of connectivity—a prediction that is supported experimentally (Baker et al., 2012; Lynch and Wagner, 2008; Stefflova et al., 2013; Tsong et al., 2006). Thus, any two regulators that overlap in their expression would be predicted to share a fraction of their targets under the conditions they are both expressed (Lynch, 2007a), unless these additional connections are specifically selected against.

Formation of Common Network Motifs

One strategy to simplify and understand the function of complex transcription networks has been to search for network motifs—

biology, and it seems likely that, in its most general form (but not necessarily in its detailed mechanism), it is often under purifying selection.

But what about motifs other than positive feedback loops? A more complex network motif known as a feed-forward loop (in which one regulator controls a second regulator and both control the same target gene) is overrepresented in biological networks (Milo et al., 2002). Depending on the parameters of binding, a given feed-forward loop can, in principle, perform logic operations such as pulse detection or expression delay (Alon, 2007; Davidson, 2010). However, it is currently unclear whether the majority of naturally occurring feed-forward loops meet the types of parameter requirements needed for these behaviors.

There are thousands of feed-forward loops embedded in the two networks of Figure 1. We note that feed-forward loops are common byproducts of the evolutionary paths diagrammed in Figures 2 and 3, and thus, the preponderance of feed-forward loops in biological networks may be a result of the same non-adaptive processes that result in large network size and interconnectedness. Indeed, it has been explicitly proposed that many network motifs have arisen as a result of neutral evolutionary processes (Cordero and Hogeweg, 2006; Ingram et al., 2006; Ruths and Nakhleh, 2013; Ward and Thornton, 2007). These ideas contrast with models in which each feed-forward loop in the network possesses optimized parameters that specify a particular transcriptional input-output relationship.

We note that feed-forward loops may also represent non-adaptive intermediates between alternative forms of transcriptional regulation (Figure 5). Rewiring of transcription networks,

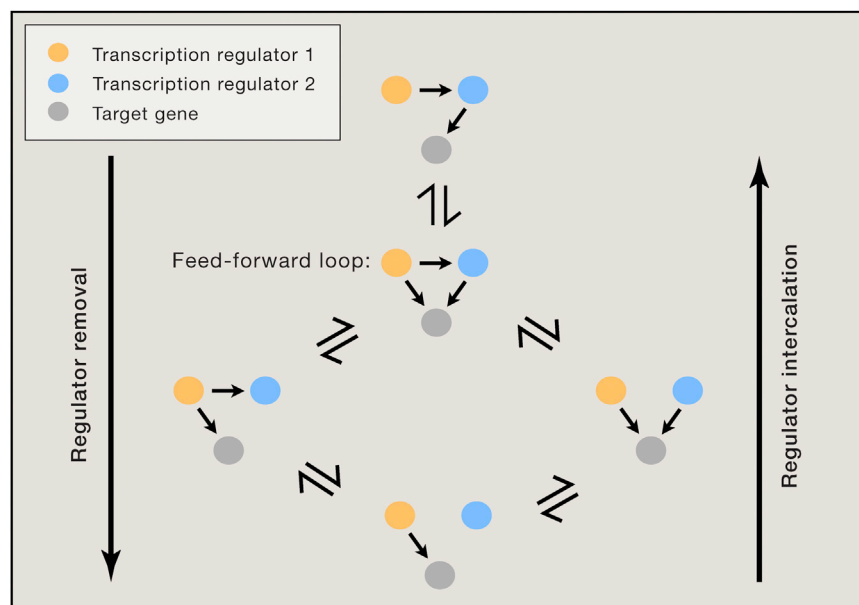


Figure 5. Pathways for Incorporation or Removal of a Transcription Regulator without Breaking the Network

Removal of the blue regulator from the linear regulatory pathway shown in the top network diagram can proceed by first forming a feed-forward loop. Subsequent loss of the connections between the orange and blue regulator and between the blue regulator and the target gene will completely remove the blue regulator from the network as shown in the bottom diagram. The opposite process starting from the bottom diagram and proceeding to the top results in intercalation of the blue regulator into the pathway. If the functions of the blue and orange regulators are redundant in the context of the network, the network can drift between these configurations over evolutionary time without compromising the output of the circuit.

at least in some cases, proceeds through intermediates that are regulated by both the ancestral and derived mechanisms (Li and Johnson, 2010), allowing the regulatory output to be preserved during the rewiring. Although they might arise non-adaptively, feed-forward loops can serve as redundant intermediates between the ancestral and derived states, and thus many observed examples of transcription network rewiring may be a simple consequence of the high frequency with which feed-forward loops are formed by neutral evolution (Lynch, 2007a).

Conclusions

Genomes evolve under selective pressure, but we no longer expect their structures to be orderly, logical affairs dictated by underlying design principles. Here, we have argued that there is no reason to expect transcription circuit networks to be any different. We argue that the drift of transcription networks to steady-state levels of high complexity and interconnection is consistent with the biochemical and biophysical properties of regulatory proteins and *cis*-regulatory sequences, particularly the cooperative binding of transcription regulators to DNA. Combined with universal processes of evolution such as mutation, genetic drift, and selection, network complexity is predicted, from first principles, to be a natural consequence. Complex structures, even if they arose in populations non-adaptively, can nonetheless serve as substrates for future evolutionary innovations or be locked in place by secondary changes; thus, they can be retained by purifying selection even though they arose non-adaptively and are not optimized solutions. If transcription circuits are considered as relatively crude products whose structures are dominated by high-probability evolutionary pathways, many of their more baffling features—size, similarity across diverse species, complexity, redundancy, interconnectedness, for example—begin to make conceptual sense.

ACKNOWLEDGMENTS

We thank C. Nobile, E. Mancera, S. Coyle, and V. Hanson-Smith for helpful comments on the manuscript. We also thank K. Carniol and two anonymous reviewers for providing critical feedback on the content and language. Work from the authors' laboratory was supported by grant R01 GM037049 from the NIH. T.R.S. was supported by a Graduate Research Fellowship from the National Science Foundation.

REFERENCES

- Alon, U. (2007). Network motifs: theory and experimental approaches. *Nat. Rev. Genet.* 8, 450–461.
- Baker, C.R., Tuch, B.B., and Johnson, A.D. (2011). Extensive DNA-binding specificity divergence of a conserved transcription regulator. *Proc. Natl. Acad. Sci. USA* 108, 7493–7498.
- Baker, C.R., Booth, L.N., Sorrells, T.R., and Johnson, A.D. (2012). Protein modularity, cooperative binding, and hybrid regulatory states underlie transcriptional network diversification. *Cell* 151, 80–95.
- Bateman, E. (1998). Autoregulation of eukaryotic transcription factors. *Prog. Nucleic Acid Res. Mol. Biol.* 60, 133–168.
- Borneman, A.R., Leigh-Bell, J.A., Yu, H., Bertone, P., Gerstein, M., and Snyder, M. (2006). Target hub proteins serve as master regulators of development in yeast. *Genes Dev.* 20, 435–448.
- Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* 122, 947–956.
- Boyle, A.P., Araya, C.L., Brdlik, C., Cayting, P., Cheng, C., Cheng, Y., Gardner, K., Hillier, L.W., Janette, J., Jiang, L., et al. (2014). Comparative analysis of regulatory information and circuits across distant species. *Nature* 512, 453–456.
- Carroll, S.B. (2008). Evo-devo and an expanding evolutionary synthesis: a genetic theory of morphological evolution. *Cell* 134, 25–36.
- Chan, S.S.-K., and Kyba, M. (2013). What is a master regulator? *J. Stem Cell Res. Ther.* Published online May 4, 2013. <http://dx.doi.org/10.4172/2157-7633.1000e114>.
- Cordero, O.X., and Hogeweg, P. (2006). Feed-forward loop circuits as a side effect of genome evolution. *Mol. Biol. Evol.* 23, 1931–1936.

- Covello, P.S., and Gray, M.W. (1993). On the evolution of RNA editing. *Trends Genet.* 9, 265–268.
- Davidson, E.H. (2010). Emerging properties of animal gene regulatory networks. *Nature* 468, 911–920.
- Doolittle, W.F. (2013). Is junk DNA bunk? A critique of ENCODE. *Proc. Natl. Acad. Sci. USA* 110, 5294–5300.
- Featherstone, D.E., and Broadie, K. (2002). Wrestling with pleiotropy: genomic and topological analysis of the yeast gene expression network. *BioEssays* 24, 267–274.
- Fischer, J.A., Giniger, E., Maniatis, T., and Ptashne, M. (1988). GAL4 activates transcription in *Drosophila*. *Nature* 332, 853–856.
- Fisher, W.W., Li, J.J., Hammonds, A.S., Brown, J.B., Pfeiffer, B.D., Weiszm, R., MacArthur, S., Thomas, S., Stamatoyannopoulos, J.A., Eisen, M.B., et al. (2012). DNA regions bound at low occupancy by transcription factors do not drive patterned reporter gene expression in *Drosophila*. *Proc. Natl. Acad. Sci. USA* 109, 21330–21335.
- Force, A., Lynch, M., Pickett, F.B., Amores, A., Yan, Y.L., and Postlethwait, J. (1999). Preservation of duplicate genes by complementary, degenerative mutations. *Genetics* 151, 1531–1545.
- Frankel, N., Wang, S., and Stern, D.L. (2012). Conserved regulatory architecture underlies parallel genetic changes and convergent phenotypic evolution. *Proc. Natl. Acad. Sci. USA* 109, 20975–20979.
- Galgoczy, D.J., Cassidy-Stone, A., Llinás, M., O'Rourke, S.M., Herskowitz, I., DeRisi, J.L., and Johnson, A.D. (2004). Genomic dissection of the cell-type-specification circuit in *Saccharomyces cerevisiae*. *Proc. Natl. Acad. Sci. USA* 101, 18069–18074.
- Gerhart, J., and Kirschner, M. (1997). *Cells, Embryos and Evolution* (Wiley-Blackwell).
- Gray, M.W., Lukes, J., Archibald, J.M., Keeling, P.J., and Doolittle, W.F. (2010). Cell biology. Irremediable complexity? *Science* 330, 920–921.
- Guelzim, N., Bottani, S., Bourguin, P., and Képès, F. (2002). Topological and causal structure of the yeast transcriptional regulatory network. *Nat. Genet.* 31, 60–63.
- Halder, G., Callaerts, P., and Gehring, W.J. (1995). Induction of ectopic eyes by targeted expression of the eyeless gene in *Drosophila*. *Science* 267, 1788–1792.
- Hayes, T.E., Sengupta, P., and Cochran, B.H. (1988). The human c-fos serum response factor and the yeast factors GRM/PRTF have related DNA-binding specificities. *Genes Dev.* 2 (12B), 1713–1722.
- Hernday, A.D., Lohse, M.B., Fordyce, P.M., Nobile, C.J., DeRisi, J.L., and Johnson, A.D. (2013). Structure of the transcriptional network controlling white-opaque switching in *Candida albicans*. *Mol. Microbiol.* 90, 22–35.
- Ingram, P.J., Stumpf, M.P.H., and Stark, J. (2006). Network motifs: structure does not determine function. *BMC Genomics* 7, 108.
- Iyer, V.R., Horak, C.E., Scafe, C.S., Botstein, D., Snyder, M., and Brown, P.O. (2001). Genomic binding sites of the yeast cell-cycle transcription factors SBF and MBF. *Nature* 409, 533–538.
- Johnson, A.D., Meyer, B.J., and Ptashne, M. (1979). Interactions between DNA-bound repressors govern regulation by the lambda phage repressor. *Proc. Natl. Acad. Sci. USA* 76, 5061–5065.
- Junion, G., Spivakov, M., Girardot, C., Braun, M., Gustafson, E.H., Birney, E., and Furlong, E.E.M. (2012). A transcription factor collective defines cardiac cell fate and reflects lineage history. *Cell* 148, 473–486.
- Kakidani, H., and Ptashne, M. (1988). GAL4 activates gene expression in mammalian cells. *Cell* 52, 161–167.
- Kielbasa, S.M., and Vingron, M. (2008). Transcriptional autoregulatory loops are highly conserved in vertebrate evolution. *PLoS ONE* 3, e3210.
- Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. (2008). An extended transcriptional network for pluripotency of embryonic stem cells. *Cell* 132, 1049–1061.
- Lavoie, H., Hogues, H., Mallick, J., Sellam, A., Nantel, A., and Whiteway, M. (2010). Evolutionary tinkering with conserved components of a transcriptional regulatory network. *PLoS Biol.* 8, e1000329.
- Lee, T.I., Rinaldi, N.J., Robert, F., Odom, D.T., Bar-Joseph, Z., Gerber, G.K., Hannett, N.M., Harbison, C.T., Thompson, C.M., Simon, I., et al. (2002). Transcriptional regulatory networks in *Saccharomyces cerevisiae*. *Science* 298, 799–804.
- Li, H., and Johnson, A.D. (2010). Evolution of transcription networks—lessons from yeasts. *Curr. Biol.* 20, R746–R753.
- Liang, Z., and Biggin, M.D. (1998). Eve and ftz regulate a wide array of genes in blastoderm embryos: the selector homeoproteins directly or indirectly regulate most genes in *Drosophila*. *Development* 125, 4471–4482.
- Lin, S., and Riggs, A.D. (1975). The general affinity of lac repressor for *E. coli* DNA: implications for gene regulation in procaryotes and eucaryotes. *Cell* 4, 107–111.
- Little, J.W. (2010). Evolution of complex gene regulatory circuits by addition of refinements. *Curr. Biol.* 20, R724–R734.
- Ludwig, M.Z., Bergman, C., Patel, N.H., and Kreitman, M. (2000). Evidence for stabilizing selection in a eukaryotic enhancer element. *Nature* 403, 564–567.
- Lukeš, J., Archibald, J.M., Keeling, P.J., Doolittle, W.F., and Gray, M.W. (2011). How a neutral evolutionary ratchet can build cellular complexity. *IUBMB Life* 63, 528–537.
- Lynch, M. (2007a). The evolution of genetic networks by non-adaptive processes. *Nat. Rev. Genet.* 8, 803–813.
- Lynch, M. (2007b). The frailty of adaptive hypotheses for the origins of organismal complexity. *Proc. Natl. Acad. Sci. USA* 104 (1), 8597–8604.
- Lynch, V.J., and Wagner, G.P. (2008). Resurrecting the role of transcription factor change in developmental evolution. *Evolution* 62, 2131–2154.
- Lynch, M., Field, M.C., Goodson, H.V., Malik, H.S., Pereira-Leal, J.B., Roos, D.S., Turkewitz, A.P., and Sazer, S. (2014). Evolutionary cell biology: two origins, one objective. *Proc. Natl. Acad. Sci. USA* 111, 16990–16994.
- MacArthur, S., Li, X.-Y., Li, J., Brown, J.B., Chu, H.C., Zeng, L., Grondona, B.P., Hechmer, A., Simirenko, L., Keränen, S.V.E., et al. (2009). Developmental roles of 21 *Drosophila* transcription factors are determined by quantitative differences in binding to an overlapping set of thousands of genomic regions. *Genome Biol.* 10, R80.
- Martchenko, M., Levitin, A., Hogues, H., Nantel, A., and Whiteway, M. (2007). Transcriptional rewiring of fungal galactose-metabolism circuitry. *Curr. Biol.* 17, 1007–1013.
- Mastick, G.S., McKay, R., Oligino, T., Donovan, K., and López, A.J. (1995). Identification of target genes regulated by homeotic proteins in *Drosophila melanogaster* through genetic selection of Ultrabithorax protein-binding sites in yeast. *Genetics* 139, 349–363.
- Milo, R., Shen-Orr, S., Itzkovitz, S., Kashtan, N., Chklovskii, D., and Alon, U. (2002). Network motifs: simple building blocks of complex networks. *Science* 298, 824–827.
- Monteiro, A. (2012). Gene regulatory networks reused to build novel traits: co-option of an eye-related gene regulatory network in eye-like organs and red wing patches on insect wings is suggested by optix expression. *BioEssays* 34, 181–186.
- Mösch, H.U., and Fink, G.R. (1997). Dissection of filamentous growth by transposon mutagenesis in *Saccharomyces cerevisiae*. *Genetics* 145, 671–684.
- Moses, A.M., Pollard, D.A., Nix, D.A., Iyer, V.N., Li, X.-Y., Biggin, M.D., and Eisen, M.B. (2006). Large-scale turnover of functional transcription factor binding sites in *Drosophila*. *PLoS Comput. Biol.* 2, e130.
- Nobile, C.J., Fox, E.P., Nett, J.E., Sorrells, T.R., Mitrovich, Q.M., Hernday, A.D., Tuch, B.B., Andes, D.R., and Johnson, A.D. (2012). A recently evolved transcriptional network controls biofilm development in *Candida albicans*. *Cell* 148, 126–138.
- Novershtern, N., Subramanian, A., Lawton, L.N., Mak, R.H., Haining, W.N., McConkey, M.E., Habib, N., Yosef, N., Chang, C.-Y., Shay, T., et al. (2011).

- Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* 144, 296–309.
- Paixão, T., and Azevedo, R.B.R. (2010). Redundancy and the evolution of cis-regulatory element multiplicity. *PLoS Comput. Biol.* 6, e1000848.
- Pavlicev, M., and Wagner, G.P. (2012). A model of developmental evolution: selection, pleiotropy and compensation. *Trends Ecol. Evol.* 27, 316–322.
- Pires, N.D., Yi, K., Breuninger, H., Catarino, B., Menand, B., and Dolan, L. (2013). Recruitment and remodeling of an ancient gene regulatory network during land plant evolution. *Proc. Natl. Acad. Sci. USA* 110, 9571–9576.
- Polach, K.J., and Widom, J. (1996). A model for the cooperative binding of eukaryotic regulatory proteins to nucleosomal target sites. *J. Mol. Biol.* 258, 800–812.
- Ptashne, M., and Gann, A. (2002). *Genes and Signals* (Cold Spring Harbor Laboratory Press).
- Qian, W., Ma, D., Xiao, C., Wang, Z., and Zhang, J. (2012). The genomic landscape and evolutionary resolution of antagonistic pleiotropy in yeast. *Cell Rep.* 2, 1399–1410.
- Raff, R.A., and Kaufman, T.C. (1983). *Embryos, Genes, and Evolution* (Macmillan Publishing Company).
- Reece-Hoyes, J.S., Pons, C., Diallo, A., Mori, A., Shrestha, S., Kadreppa, S., Nelson, J., Diprima, S., Dricot, A., Lajoie, B.R., et al. (2013). Extensive rewiring and complex evolutionary dynamics in a *C. elegans* multiparameter transcription factor network. *Mol. Cell* 51, 116–127.
- Reed, R.D., Papa, R., Martin, A., Hines, H.M., Counterman, B.A., Pardo-Diaz, C., Jiggins, C.D., Chamberlain, N.L., Kronforst, M.R., Chen, R., et al. (2011). *optix* drives the repeated convergent evolution of butterfly wing pattern mimicry. *Science* 333, 1137–1141.
- Ren, B., Robert, F., Wyrick, J.J., Aparicio, O., Jennings, E.G., Simon, I., Zeitlinger, J., Schreiber, J., Hannett, N., Kanin, E., et al. (2000). Genome-wide location and function of DNA binding proteins. *Science* 290, 2306–2309.
- Ruths, T., and Nakhleh, L. (2012). ncDNA and drift drive binding site accumulation. *BMC Evol. Biol.* 12, 159.
- Ruths, T., and Nakhleh, L. (2013). Neutral forces acting on intragenomic variability shape the *Escherichia coli* regulatory network topology. *Proc. Natl. Acad. Sci. USA* 110, 7754–7759.
- Sayou, C., Monniaux, M., Nanao, M.H., Moyroud, E., Brockington, S.F., Thévenon, E., Chahtane, H., Warthmann, N., Melkonian, M., Zhang, Y., et al. (2014). A promiscuous intermediate underlies the evolution of LEAFY DNA binding specificity. *Science* 343, 645–648.
- Schmidt, D., Wilson, M.D., Ballester, B., Schwalie, P.C., Brown, G.D., Marshall, A., Kutter, C., Watt, S., Martinez-Jimenez, C.P., Mackay, S., et al. (2010). Five-vertebrate ChIP-seq reveals the evolutionary dynamics of transcription factor binding. *Science* 328, 1036–1040.
- Schweizer, A., Rupp, S., Taylor, B.N., Rölinghoff, M., and Schröppel, K. (2000). The TEA/ATTS transcription factor CaTec1p regulates hyphal development and virulence in *Candida albicans*. *Mol. Microbiol.* 38, 435–445.
- Stefflova, K., Thybert, D., Wilson, M.D., Streeter, I., Aleksic, J., Karagianni, P., Brazma, A., Adams, D.J., Talianidis, I., Marioni, J.C., et al. (2013). Cooperativity and rapid evolution of cobound transcription factors in closely related mammals. *Cell* 154, 530–540.
- Stergachis, A.B., Neph, S., Sandstrom, R., Haugen, E., Reynolds, A.P., Zhang, M., Byron, R., Canfield, T., Stelting-Sun, S., Lee, K., et al. (2014). Conservation of trans-acting circuitry during mammalian regulatory evolution. *Nature* 515, 365–370.
- Stern, D.L., and Orgogozo, V. (2009). Is genetic evolution predictable? *Science* 323, 746–751.
- Stoltzfus, A. (1999). On the possibility of constructive neutral evolution. *J. Mol. Evol.* 49, 169–181.
- Stone, J.R., and Wray, G.A. (2001). Rapid evolution of cis-regulatory sequences via local point mutations. *Mol. Biol. Evol.* 18, 1764–1770.
- Sullivan, A.M., Arsovski, A.A., Lempe, J., Bubb, K.L., Weirauch, M.T., Sabo, P.J., Sandstrom, R., Thurman, R.E., Neph, S., Reynolds, A.P., et al. (2014). Mapping and dynamics of regulatory DNA and transcription factor networks in *A. thaliana*. *Cell Rep.* 8, 2015–2030.
- Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* 126, 663–676.
- Tanay, A., Regev, A., and Shamir, R. (2005). Conservation and evolvability in regulatory networks: the evolution of ribosomal regulation in yeast. *Proc. Natl. Acad. Sci. USA* 102, 7203–7208.
- Tapscott, S.J., Davis, R.L., Thayer, M.J., Cheng, P.F., Weintraub, H., and Lasar, A.B. (1988). MyoD1: a nuclear phosphoprotein requiring a Myc homology region to convert fibroblasts to myoblasts. *Science* 242, 405–411.
- Tsong, A.E., Miller, M.G., Raisner, R.M., and Johnson, A.D. (2003). Evolution of a combinatorial transcriptional circuit: a case study in yeasts. *Cell* 115, 389–399.
- Tsong, A.E., Tuch, B.B., Li, H., and Johnson, A.D. (2006). Evolution of alternative transcriptional circuits with identical logic. *Nature* 443, 415–420.
- Tuch, B.B., Galgoczy, D.J., Hernday, A.D., Li, H., and Johnson, A.D. (2008a). The evolution of combinatorial gene regulation in fungi. *PLoS Biol.* 6, e38.
- Tuch, B.B., Li, H., and Johnson, A.D. (2008b). Evolution of eukaryotic transcription circuits. *Science* 319, 1797–1799.
- Tursun, B., Patel, T., Kratsios, P., and Hobert, O. (2011). Direct conversion of *C. elegans* germ cells into specific neuron types. *Science* 331, 304–308.
- Vierbuchen, T., Ostermeier, A., Pang, Z.P., Kokubu, Y., Südhof, T.C., and Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* 463, 1035–1041.
- Wagner, A. (2003). Does selection mold molecular networks? *Sci. STKE* 2003, PE41.
- Wang, D.Y.C., Kumar, S., and Hedges, S.B. (1999). Divergence time estimates for the early history of animal phyla and the origin of plants, animals and fungi. *Proc. Biol. Sci.* 266, 163–171.
- Ward, J.J., and Thornton, J.M. (2007). Evolutionary models for formation of network motifs and modularity in the *Saccharomyces* transcription factor network. *PLoS Comput. Biol.* 3, 1993–2002.
- Weirauch, M.T., and Hughes, T.R. (2010). Conserved expression without conserved regulatory sequence: the more things change, the more they stay the same. *Trends Genet.* 26, 66–74.
- Weirauch, M.T., and Hughes, T.R. (2011). A catalogue of eukaryotic transcription factor types, their evolutionary origin, and species distribution. *Subcell. Biochem.* 52, 25–73.
- Whitfield, T.W., Wang, J., Collins, P.J., Partridge, E.C., Aldred, S.F., Trinklein, N.D., Myers, R.M., and Weng, Z. (2012). Functional analysis of transcription factor binding sites in human promoters. *Genome Biol.* 13, R50.
- Wray, G.A., Hahn, M.W., Abouheif, E., Balhoff, J.P., Pizer, M., Rockman, M.V., and Romano, L.A. (2003). The evolution of transcriptional regulation in eukaryotes. *Mol. Biol. Evol.* 20, 1377–1419.
- Zordan, R.E., Miller, M.G., Galgoczy, D.J., Tuch, B.B., and Johnson, A.D. (2007). Interlocking transcriptional feedback loops control white-opaque switching in *Candida albicans*. *PLoS Biol.* 5, e256.
- Zuckerandl, E. (1997). Neutral and nonneutral mutations: the creative mix—evolution of complexity in gene interaction systems. *J. Mol. Evol.* 44 (1), S2–S8.

Genesis of Chromatin and Transcription Dynamics in the Origin of Species

Maria J.E. Koster,¹ Berend Snel,² and H.Th. Marc Timmers^{1,*}

¹Molecular Cancer Research, Center for Molecular Medicine, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, The Netherlands

²Theoretical Biology and Bioinformatics, Utrecht University, Padualaan 8, 3584 CH Utrecht, The Netherlands

*Correspondence: h.t.m.timmers@umcutrecht.nl

<http://dx.doi.org/10.1016/j.cell.2015.04.033>

Histone proteins compact and stabilize the genomes of Eukarya and Archaea. By forming nucleosome(-like) structures they restrict access of DNA-binding transcription regulators to *cis*-regulatory DNA elements. Dynamic competition between histones and transcription factors is facilitated by different classes of proteins including ATP-dependent remodeling enzymes that control assembly, access, and editing of chromatin. Here, we summarize the knowledge on dynamics underlying transcriptional regulation across the domains of life with a focus on ATP-dependent enzymes in chromatin structure or in TATA-binding protein activity. These insights suggest directions for future studies on the evolution of transcription regulation and chromatin dynamics.

“Nothing in biology makes sense except in the light of evolution” is the title of an influential essay (Dobzhansky, 1973), which appeared in 1973 from the hand of the famous geneticist Theodosius Dobzhansky to empower American teachers for the creation-evolution debate in their class rooms. As a comparative zoologist, Dobzhansky was fascinated by the diversity of species. Nevertheless, he was well aware that the unity of life resides in “biochemical universals” like DNA, RNA, proteins, and certain metabolites. How could Dobzhansky know that around the time of his writing Fred Sanger was developing a rapid method for sequencing DNA (Sanger et al., 1977) and that “Sanger” sequencing of genomic DNA was about to transform his comparative zoology into comparative genomics?

Different branches of the tree of life developed distinct strategies to accurately express their genes. With increased genome size and biological complexity comes an increase in complexity of gene regulation mechanisms. The most pervasive is regulation at transcription initiation, which will be the focal point for our discussions. Transcriptional pausing is a later evolutionary invention, and excellent reviews on this appeared recently (Adelman and Lis, 2012; Yamaguchi et al., 2013). Here, we discuss the molecular functions and genomic occurrences of key components of the DNA transcription machinery across the archaeal and eukaryotic lineages in light of “adaptive” gene expression and transcriptional dynamics. In particular, we focus on evolutionary retention and expansion of the class of ATP-dependent enzymes, which are relevant for gene transcription by mammalian RNA polymerase II (pol II) and control the dynamics of chromatin structures or of basal transcription complexes. In the spirit of Dobzhansky, we aim to understand the dynamics of transcriptional regulation from an evolutionary perspective.

Transcriptional Mechanisms across the Domains of Life

The regulated action of DNA-dependent RNA polymerases in gene transcription underlies all life processes. Early studies on

adaptive gene expression in the colon bacterium, *Escherichia coli*, and its bacteriophage λ (Jacob and Monod, 1961; Ptashne, 2005) revealed that facilitated access of RNA polymerase (RNAP) to promoter DNA is regulatory for gene expression. This paradigm proved valid for all Bacteria and is also central in understanding of gene regulation in Archaea and Eukarya (Jun et al., 2011; Ptashne, 2005; Struhl, 1999). Whereas archaeal transcription units are typically of an operon-type and archaeal gene-specific regulators preclude or enhance promoter binding of RNAP and its associated factors via direct interactions, the archaeal basal transcription machinery is more similar to eukaryotic than to bacterial systems (Figure 1) (Grohmann and Werner, 2011; Jun et al., 2011). Orthologs of the basal transcription factors TATA-binding protein (TBP) and TFIIB (called TFB) direct promoter recruitment of archaeal RNAP, whereas bacterial RNAP requires a single specificity (σ) factor for promoter recognition (Grohmann and Werner, 2011; Jun et al., 2011). It was proposed that analogous to bacterial σ -factors, different combinations of TBP/TFB paralogs could be used for subsets of genes in Archaea (Grohmann and Werner, 2011; Jun et al., 2011). However, little proof for this attractive model has been obtained and a significant functional redundancy may exist between archaeal TBP and TFB paralogs (Santangelo et al., 2007). In addition, the histone proteins essential for packing chromatin into the eukaryotic nucleus are present in some Archaea (Malik and Henikoff, 2003; Reeve, 2003). It was recently proposed that the eukaryotic nuclear lineage potentially originated within present-day Archaea (Williams et al., 2013). In contrast to archaeal organisms, eukaryotes contain three RNA polymerases for the transcription of nuclear genes, which are the result of a massive “big-bang” of gene duplications during the transition from an archaeum to a fully fledged eukaryote (Koonin, 2007). Each eukaryotic RNA polymerase has a dedicated set of transcription initiation factors, which recruit the enzymes and assist in formation of the open complex competent for transcription initiation.

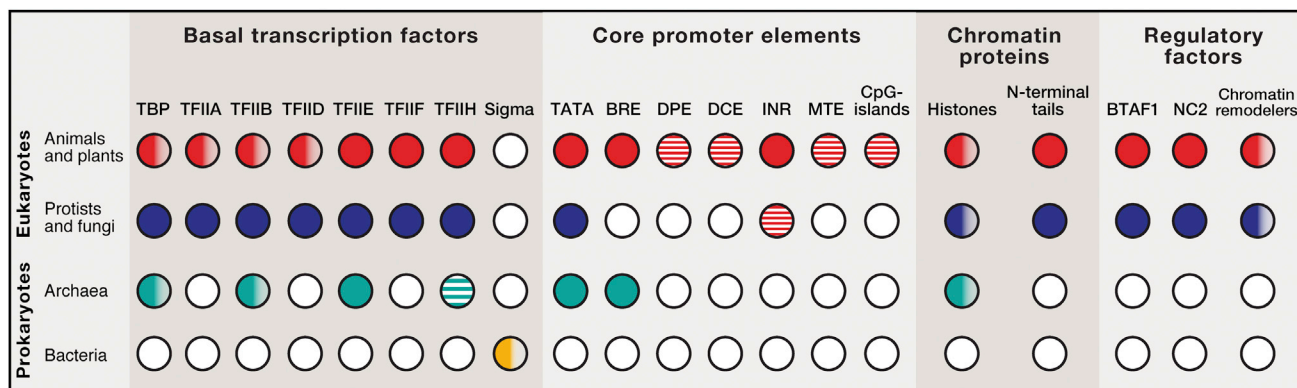


Figure 1. The Evolution of Gene Transcription

Simplified overview of the evolutionary conservation and diversification of factors important for gene transcription in the domains of life: Archaea, Bacteria, and Eukarya (split into two groups: Animals and Plants and Protists and Fungi). Filled bullets indicate orthologous proteins or sequences in the whole lineage, striped bullets the presence in part of the lineage, and open bullets that no homologs have been found. Gradient colors denote presence of paralogs. Although TFIH subunits are present in Archaea, their role is probably restricted to DNA repair. Please note that in *S. cerevisiae* the TATA and INR elements are at variable distance and it is unclear whether yeast TFIID directly contacts the INR. Also, several members of the Apicomplexa lineage lack TAFs and basal transcription factor genes (see text).

However, each RNA polymerase initiation system depends on TBP and TFIIB paralogs (Akhtar and Veenstra, 2011; Gazdag et al., 2007; Vannini and Cramer, 2012). Of all RNA synthesis machineries, eukaryotic pol II is the most versatile as it serves the largest diversity of gene promoters. It is also the most tightly controlled serving the widest dynamic range of RNA expression levels (Levine et al., 2014).

Control and Dynamics of RNA Polymerase II-Mediated Transcription

Transcription initiation by pol II is controlled roughly at three different levels. First, gene-specificity is achieved through DNA-sequence-specific binding by activator and repressor proteins (gene-specific transcription factors [GSTFs]), which serve to mark a gene promoter or enhancer for activity (Figure 2). In general, GSTF binding to DNA is highly dynamic with in vivo residence times ranging from milliseconds to a few minutes (Chen et al., 2014; Dinant et al., 2009; Hager et al., 2009). This corresponds well with FRAP (fluorescence recovery after photobleaching) experiments indicating that diffusion is the prime means for GSTF movement in living cells. It allows a GSTF to scan the entire volume of a mammalian nucleus in a matter of minutes (Hager et al., 2009). GSTF binding to its DNA target sequence in chromatin is mostly transient, but exceptions exist like yeast Rap1p and activated *Drosophila* HSF (Lickwar et al., 2012; Yao et al., 2006). DNA residence time is correlated with transcriptional output as slower exchanges correlate with increased mRNA levels (Lickwar et al., 2012; Stavreva et al., 2004).

The second level of control is exerted by transcriptional co-activator and co-repressor complexes, which often act through chromatin structures and modifications (Figure 2). These complexes are recruited to specific genomic elements by GSTFs, by chromatin modifications, by DNA, and in some cases by regulatory RNAs. While genomic binding of these chromatin-regulatory complexes is dynamic (Hager et al., 2009; Johnson et al.,

2008), their effect on chromatin function can be lasting due to the immobile character of histones and the DNA fiber in the eukaryotic nucleus (Kimura and Cook, 2001). Archaea seem to lack chromatin-remodeling and -modifying complexes, but most archaeal species contain histone-type or nucleoid proteins (Figure 1) (Sandman and Reeve, 2005). Archaeal histones are also characterized by a histone-fold domain (HFD) comprised of three α helices, but they lack the extensively modified extensions of their eukaryotic counterparts (Jun et al., 2011; Malik and Henikoff, 2003). Archaeal histones are more similar to the eukaryotic histones H3/H4 than the H2A/H2B pair (Malik and Henikoff, 2003; Sandman and Reeve, 2005). Nuclease digestion of archaeal chromatin indicates that DNA follows a spiral path on the surface of multimeric histone dimer cores with a periodicity of 30 or 60 bp (Ammar et al., 2012; Maruyama et al., 2013). Archaeal chromatin proteins seem to increase the melting temperature of DNA (Reeve, 2003), and it is tempting to speculate that a prime function of archaeal histone proteins has been to protect DNA from thermal denaturation. Several archaeal lineages like hyperthermophilic Crenarchaea lack histones but instead contain other chromatin proteins like Alba, which may perform similar functions (Sandman and Reeve, 2005). Eukaryotic histones are derived from an ancestor shared with Archaea, which duplicated the histone-fold to form a “doublet histone” (Malik and Henikoff, 2003). The ancestral gene split into histone H3 and H4 to form a H3-H4 tetramer, and after duplication it also diverged into the histone H2A-H2B heterodimer. While histone H4 is remarkably conserved, variants of H3, H2A, and H2B appeared to allow functional specialization (Malik and Henikoff, 2003). The nucleosomal repeating unit of eukaryotic chromatin consists of two copies of histone H3, H4, H2A, and H2B wrapping ~150 bp of DNA in 1.7 left-handed turns (Luger et al., 1997). Depending on linker length, nucleosomes can form higher-order structures with di-nucleosomes in head-to-head arrangements (Song et al., 2014). Eukaryotic chromatin is inherently stable and has been proposed to “maintain the restrictive

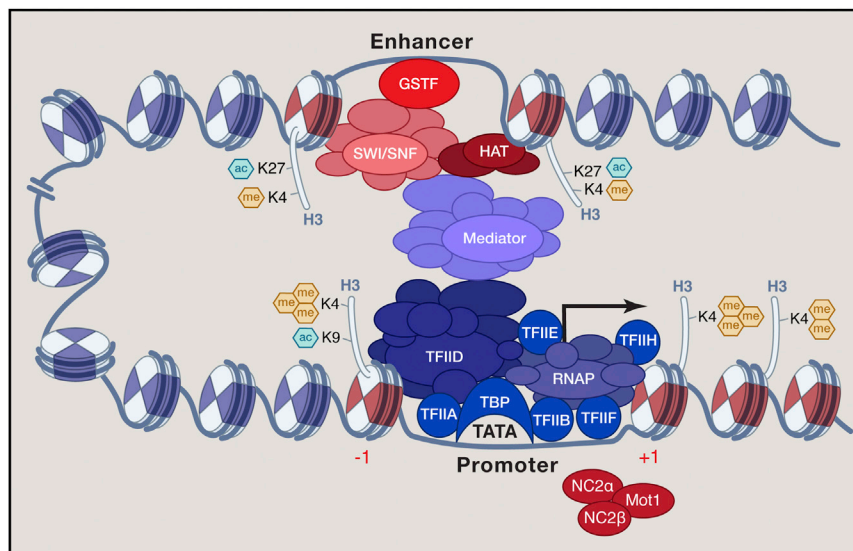


Figure 2. The Control of Transcription Initiation and Dynamics

Gene-specific transcription factors (GSTFs) bind to DNA elements to recruit regulatory complexes such as Mediator, histone acetyltransferases, and chromatin remodelers (SWI/SNF) altering chromatin structure. Pre-initiation complex (PIC) assembly starts with binding of TFIID, including TATA-binding protein (TBP) to the core promoter. Promoter association of TFIID is stabilized by TBP-associated factors (TAFs) binding to (dynamically) modified histone tails. BTA1/Mot1p and NC2 can remove TBP from the promoter. Intrinsically mobile proteins are indicated in red, while the more stably bound are colored blue.

ground state of promoters by blocking association of the basal pol II machinery with the core promoter, while permitting many GSTFs to bind their target sites" (Struhl, 1999). Interestingly, transcription regulatory regions display a paucity of nucleosomes (see below). In contrast, archaeal chromatin is relatively flexible and unstable, which allows its promoters to be accessible (Reeve, 2003; Sandman and Reeve, 2005).

The third level is formed by the pol II pre-initiation complex (PIC), which besides pol II itself consists of six basal (or general) transcription factors (Thomas and Chiang, 2006; Vannini and Cramer, 2012). PIC assembly *in vitro* is sequential (Buratowski et al., 1989) and starts with core promoter binding by the TFIID complex, which consists of TBP and 13 highly conserved TAFs (TBP-associated factors) (Papai et al., 2011). While TBP *in vitro* directly recognizes the TATA box, promoter binding by TFIID can be stabilized by binding of TAFs to core promoter DNA sequences, like the INR, DPE, MTE, and DCE (Juven-Gershon and Kadonaga, 2010) and/or binding to acetylated and methylated histone tails (Jacobson et al., 2000; Vermeulen et al., 2007). TFIID binding is stabilized by TFIIA and subsequently the remaining four basal transcription factors (TFIIB, TFIIF, TFIIE, TFIIH) and pol II itself enter to complete PIC assembly (Figure 2). It is important to note that TAFs are eukaryotic inventions, which are lacking from Archaea. The occurrence of core promoter sequences other than TATA and INR differs between eukaryotic species (Figure 1). Most of mammalian promoters reside in CpG-islands and lack a canonical TATA box (Sandelin et al., 2007). The combination of nucleosome depleted regions (NDRs), core promoter sequence elements and histone tail interactions positions TFIID onto mammalian core promoters (Cler et al., 2009; Lauberth et al., 2013; Müller and Tora, 2014; Vermeulen et al., 2007). It is interesting to note that yeast pol II promoters contain AT-rich sequences and that TATA-less promoters predominate in larger genomes (Juven-Gershon and Kadonaga, 2010; Rhee and Pugh, 2012; Tora and Timmers, 2010). Biochemical experiments indicate that binding of eukaryotic TBP to TATA occurs in a linear three-step pathway resulting in severe DNA

bending (Delgadillo et al., 2009). Minor groove deformation results from insertion of two pairs of phenylalanines between the first and last di-nucleotides of the TATA box, which is compensated by a $\sim 90^\circ$ bend in promoter (Delgadillo et al.,

2009). While TBP binding is rapid, TATA box complexes with eukaryotic TBP or TFIID are long-lived (30–45 min) *in vitro* (Hoopes et al., 1998; Timmers and Sharp, 1991; Workman and Roeder, 1987). Whereas nucleosomes can obstruct TFIID binding, the opposite is also true as template pre-incubation with TFIID or TBP renders promoter activity resistant to nucleosome repression (Meisterernst et al., 1990; Workman and Roeder, 1987). This competition also seems to occur in living cells (Tirosh and Barkai, 2008; van Werven et al., 2009). Besides nucleosome organization and interaction, the assembly rate of the pol II PIC is influenced by the combination of core-promoter elements and by PIC composition (Levine et al., 2014; Sikorski and Buratowski, 2009). It is interesting to note that live-cell imaging showed that only a few of the promoter-binding events of pol II are productive (Darzacq et al., 2007). At present, much less is known of the archaeal PIC. While TATA-interaction of archaeal TBP also results in bent DNA, this interaction is extremely dynamic with on-off rates in the (sub)second range (Gietl et al., 2014). Similarly to eukaryotes, archaeal TFB can stabilize TBP/promoter interactions, but TFIIA orthologs are absent (Figure 1).

In conclusion, the general mechanisms of transcriptional regulation display similarities in organisms from the distinct domains of life and the major differences relate to fine-tuning and the dynamic behavior of chromatin structures and of TBP/TFIID complexes.

Sources of Stochastic Gene Transcription

Gene transcription has to be dynamic to meet changing environmental and cell-intrinsic demands. A highly relevant aspect for dynamic gene regulation is the stochastic nature of pol II-mediated transcription, which relies on (in)stability of DNA-transcription factor complexes (Hager et al., 2009; Munsky et al., 2012). Analysis and modeling of mRNA abundance on a single-cell basis indicated that mRNAs from constitutively expressed genes follow Poisson distributions. In contrast, regulated mRNAs follow a two-state model, in which the promoter frequently alternates between active and inactive states (Munsky et al., 2012; van

Werven et al., 2008). This behavior increases gene expression “noise,” which may enable rapid differential cellular responses to cell-external and -internal cues (Newman et al., 2006). Mutational analyses of the regulated *GAL1* promoter from yeast revealed that mutations in its canonical TATA box reduce transcriptional bursting and cell-to-cell variability in expression (Blake et al., 2006). Interestingly, genome-wide analysis in yeast cells showed that TBP turnover is higher at TATA-containing promoters compared to promoters lacking a canonical TATA box (van Werven et al., 2009). FRAP experiments in human and yeast cells indicated that TBP exists in (at least) two pools of different mobility (de Graaf et al., 2010; Sprouse et al., 2008). Another attribute of regulated promoters is that the TATA box is often occluded by nucleosomes (Tirosh and Barkai, 2008) and that assembly of a functional PIC requires (transient) removal of this +1 nucleosome. Interestingly, the TATA box is enriched in (developmentally) regulated promoters from yeast or humans (Basehoar et al., 2004; Sandelin et al., 2007). With the bulk of histone proteins being immobile in vivo (Kimura and Cook, 2001), remodeling of nucleosome structures at DNaseI-hypersensitive sites (DHSs) like promoters and enhancers, is a continuous process in cells (Hager et al., 2009). Also, histone H3 turnover analysis in yeast showed that this histone is replaced more rapidly at promoters than at coding regions and that H3 turnover rate in coding regions correlates with pol II density (Dion et al., 2007).

Together, this indicates that the biochemical stabilities of the eukaryotic histone/DNA and TBP/TATA box complexes are countered in vivo by specific dynamic processes. This may contribute to stochastic and transient promoter activation and to transcriptional noise of pol II-transcribed genes.

Moving the Immobile to meet Dynamic Demands Requires Energy

The molecular processes responsible for chromatin and PIC dynamics remained elusive until April 1992, when *Molecular and Cellular Biology* published two landmark studies identifying the yeast transcription regulators *SNF2* and *MOT1* as ATP-consuming enzymes (Davis et al., 1992; Laurent et al., 1992).

Mutations in the *SNF2* gene had been isolated in genetic screens for loss of growth on sucrose by Carlson and coworkers (Neigeborn and Carlson, 1984). Suppressor analyses of *snf2* alleles provided links with histone proteins and chromatin regulation (Hirschhorn et al., 1992). Soon after, biochemical studies showed that binding of the Gal4p GSTF to nucleosomal DNA was stimulated by a Snf2p-containing complex in an ATP-hydrolysis dependent manner (Côté et al., 1994; Kwon et al., 1994). It rapidly became clear that *SNF2* is identical to *SWI2*, which had been isolated in screens for defective mating-type switching (Stern et al., 1984). The Swi2p/Snf2p ATPase became the progenitor of the family of ATP-dependent chromatin remodelers (Clapier and Cairns, 2009). In metazoans, *SWI2/SNF2* orthologs play important roles in development and cancer (Hargreaves and Crabtree, 2011; Shain and Pollack, 2013).

Using a genetic screen for regulators of basal activity of the yeast *CYC1* promoter the Thorner group isolated *MOT1* alleles to realize that its gene product contains a helicase domain, which is homologous to Snf2p and Rad54p (Davis et al., 1992). Subsequently, Auble and Hahn showed that Mot1p binds with

high affinity to TBP-TATA complexes in vitro and uses ATP-hydrolysis to dissociate the complex (Auble and Hahn, 1993; Auble et al., 1994). Stable Mot1p-TBP complexes were isolated from yeast cell extracts (Poon et al., 1994). Parallel work with human cells showed that the orthologous complex, B-TFIID (BTAF1/hMot1p plus human TBP), can replace TFIID and TBP in in vitro transcription assays. B-TFIID rapidly exchanges between TATA boxes and contains a potent (d)ATPase activity (Timmers et al., 1992; Timmers and Sharp, 1991).

Identification of *SNF2* and *MOT1* as ATP-dependent remodelers opened studies toward the dynamics of inherently stable nucleosomal and TBP/TATA complexes. While first classified as a *SWI2/SNF2* family member, phylogenetic comparisons indicate that *MOT1* and its human ortholog *BTAF1* belong to a separate lineage within the SNF2 family of ATPases (Eisen et al., 1995; Flaus et al., 2006). This lineage includes the RAD54 ATPase involved in DNA repair and RAD54 orthologs, RAD54L2 and ATRX/RAD54L. Interestingly, the eukaryotic BTAF1/RAD54 lineage relates to a different archaeal homolog than *SWI2/SNF2* (Figure 3). In the following sections we discuss function, evolutionary retention, and expansion of gene families encoding ATP-dependent enzymes relevant for transcription and chromatin regulation in the context of their substrates. In this discussion our viewpoint will be the human genome.

Chromatin Remodelers Move and Restructure Nucleosomes

The *SWI2/SNF2* gene family expanded to 27 members in humans (Hargreaves and Crabtree, 2011). Of this family 16 of the ATP-dependent remodelers are currently implicated in controlling chromatin structures relevant for pol II transcription. The catalytic domain of all remodelers consists of two RecA-like lobes and is highly similar to that of DNA translocases (Becker and Workman, 2013). Recent models indicate that ATP-dependent remodelers employ a DNA translocation mechanism to modify chromatin structure (Bartholomew, 2014; Clapier and Cairns, 2009; Narlikar et al., 2013). Depending on the ATPase and its associated proteins, the action can be chromatin assembly, accessibility, and/or editing. Transcription-relevant remodelers have been divided into four distinct families (*SWI/SNF*, *ISWI/SNF2L*, *CHD/Mi-2*, *INO80*), which are genetically and functionally non-redundant, and we restrict our discussion to these groups (Bartholomew, 2014; Clapier and Cairns, 2009; Hargreaves and Crabtree, 2011). The combined cellular abundance of remodelers is estimated to be of one remodeling complex per four nucleosome substrates (Moshkin et al., 2012), suggesting that chromatin remodeling is a continuous process. Mutations in several remodelers or their associated subunits are causative to defects in metazoan development and to cancer in human cells, which underscores the importance of chromatin remodeling for biological processes (Hargreaves and Crabtree, 2011; Kadoch et al., 2013; Shain and Pollack, 2013). Below, we shortly describe the four distinct families, and we refer to excellent reviews with more details (Becker and Workman, 2013; Clapier and Cairns, 2009; Hargreaves and Crabtree, 2011; Narlikar et al., 2013).

SWI/SNF Group

The mammalian orthologs of yeast *SWI/SNF2* are Brg1 (*SMARCA4*) and Brm (*SMARCA2*), which form BAF complexes

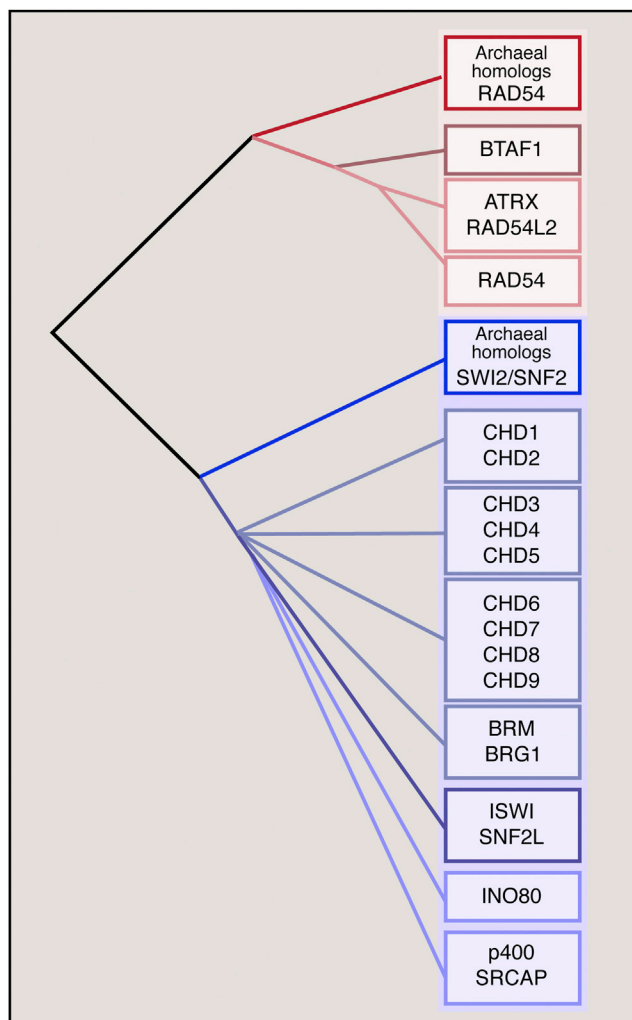


Figure 3. The Evolution of ATP-Dependent Enzymes in Transcription and Chromatin Regulation

Schematic representation of the tree of life for ATPase subunits representing the origin of the BTAF1/Mot1p-ATRX-RAD54 and CHD-SNF2-INO80-SWR lineages. The colors represent two groups that duplicated and diverged early in an archaeal and eukaryotic ancestor.

with either Brm or Brg1. Brg1 is also the catalytic subunit of the PBAF remodeler. The ATPase domain of *SWI2/SNF2* orthologs is abutted by an upstream HSA domain and a C-terminal bromo-domain, which can bind to acetylated lysines. Both in mammals and yeast, *SWI2/SNF2* proteins assemble into large remodeling complexes. Whereas only one *SWI2/SNF2* isoform is present in yeast, mammalian BAF complexes can differ in subunit make-up. Subunit exchange is used to regulate specific gene expression programs during development. Besides *SWI2/SNF2*, budding yeast also contains the RSC complex and mammalian PBAF is presumed to be the counterpart of yeast RSC (Bartholomew, 2014; Clapier and Cairns, 2009; Hargreaves and Crabtree, 2011). Recent evidence indicates that RSC action rather than AT richness is responsible for nucleosome depletion from intergenic regions in yeast (Lorch et al., 2014).

ISWI/SNF2L Group

ISWI/SNF2L remodelers also form distinct functional complexes by assembling with different homologous subunits (Hargreaves and Crabtree, 2011). ATPases of this group are characterized by SANT and SLIDE domains at their C terminus, which form a nucleosome recognition module (Clapier and Cairns, 2009). ISWI/SNF2L proteins assemble in different complexes with one to four subunits. The ISWI/SNF2L family is involved in repression of non-coding RNA transcription, heterochromatin formation, DNA replication, and ES cell pluripotency (Hargreaves and Crabtree, 2011; Koster et al., 2014). Interestingly, the fission yeast *Schizosaccharomyces pombe* lacks any ISWI/SNF2L ortholog (Pointner et al., 2012).

CHD/Mi-2 Group

Defining features for this group are two tandemly arranged chromo-domains, which lie N-terminal to the ATPase domain. Chromo-domains can interact with methylated histones and/or DNA (Clapier and Cairns, 2009). A single *CHD1* gene is present in *Saccharomyces cerevisiae* and the fission yeast genome contains three paralogs (Pointner et al., 2012), which may compensate for the absence of ISWI/SNF2L orthologs. CHD remodelers have expanded during evolution. Nine CHD genes are present in mammalian genomes, divided over three subfamilies. The first subfamily consists of CHD1 and CHD2, which contain a C-terminal DNA-binding domain. The second subfamily includes the PHD finger-containing CHD3 and CHD4. The third group is more diverse and consists of CHD5-CHD9, which have additional functional domains. Overall, the CHD/Mi-2 family is very versatile, and its members promote or repress transcription and participate in other events like mRNA processing (Murawska and Brehm, 2011).

INO80 Group

This group consists of three members in humans: INO80, SRCAP, and p400. These enzymes are characterized with a large insertion between the RecA-like lobes (Clapier and Cairns, 2009; Hargreaves and Crabtree, 2011). They form large assemblies with 8–14 subunits. SRCAP and p400 complexes exchange histone H2A.Z/H2B dimers for canonical H2A/H2B, and the INO80 complex performs the reverse reaction. The yeast SWR1 exchanger may collaborate with RSC to deposit H2A.Z/H2B abutting NDRs (Ranjan et al., 2013). An evolutionary conserved function of INO80 family members is chromatin editing. Furthermore, INO80 enzymes have been implicated in DNA repair and replication (Clapier and Cairns, 2009; Hargreaves and Crabtree, 2011).

The mechanism, which couples ATP hydrolysis to nucleosome translocation, is not well understood (Bartholomew, 2014; Narlikar et al., 2013). Various models have been proposed: “the twist diffusion” model, the “loop propagation” model, and the “octamer swiveling” model. A recent single-molecule FRET (fluorescence resonance energy transfer) study suggests the following model for nucleosome remodeling by ISWI/SNF2L enzymes: DNA is first translocated in single-bp steps toward the nucleosomal exit side by the ATPase domain; this generates strain on the entry-side DNA; after translocation of seven bps, this triggers DNA at the nucleosomal entry side to be drawn into the nucleosome; an additional three bps of DNA is translocated to the exit side; this step repeats to generate processive DNA translocation

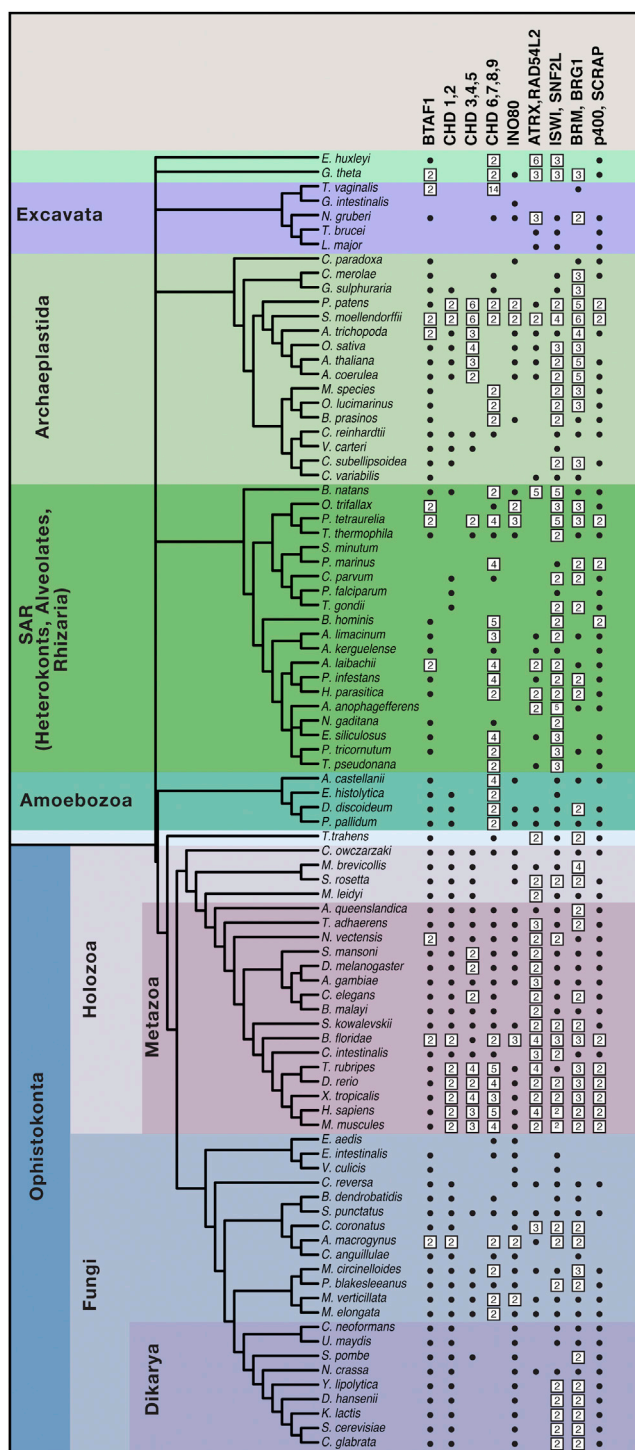


Figure 4. The Evolution of the SWI2/SNF2 Family

Schematic representation of the tree of life with a selection of eukaryotic species from the different supergroups (Excavata; Archaeplastida; SAR; Amoebozoa; Ophisthokonta) indicated on the left. The SWI2/SNF2 family member proteins are organized in different functional groups (BTAF1; CHD1,2; CHD3,4,5; CHD6,7,8,9; INO80; ATRX; RAD54L2; SNF2H,SNF2L; BRG1,BRM; SWR1), and whenever present, the number of homologs is indicated in black boxes. A filled bullet indicates presence of a single ortholog.

across the nucleosome (Bartholomew, 2014; Narlikar et al., 2013). It is likely but currently unclear whether different remodeler families utilize distinct mechanisms.

Chromatin Remodeling Complexes: Phylogenetics, Function, and Regulation

Expansion of eukaryotic genomes mandated more extensive DNA condensation and this provided evolutionary pressure to expand the chromatin-remodeling class of enzymes early on. While the catalytic domain of SWI2/SNF2 ATPases seems to have a bacterial ancestor, these domains are equipped with chromatin-binding domains in eukaryotes (Eisen et al., 1995; Flaus et al., 2006; Iyer et al., 2008). We performed phylogenetic comparisons to infer the evolutionary history of gene families encoding ATP-dependent enzymes relevant for eukaryotic transcription regulation (Figure 4). The universality of the chromatin-remodeler families supports their origin soon after the onset of the eukaryotic lineage but before the initial radiation of eukaryotic species. Higher eukaryotes further expanded the number of genes encoding these ATPases and associated subunits through gene duplication (Hargreaves and Crabtree, 2011). Together with the acquisition of novel domains, proliferation of paralogous families led to a diverse set of enzymes. Most eukaryotes have representatives of all four classes of remodelers (SWI/SNF, ISWI/SNF2L, CHD/Mi-2, INO80). This allows higher eukaryotes to form distinct functional complexes that drive and maintain developmental and cell-type-specific gene expression programs. The early divergence and in some cases duplication of plant homologs resulted in plant-specific chromatin remodelers with functions deviating from their metazoan counterparts (Gentry and Hennig, 2014).

Our current knowledge of chromatin-based mechanisms controlling transcriptional permissiveness is derived from a limited set of protozoan and metazoan model organisms, which may not represent the full spectrum. For example, the coral symbiont and dinoflagellate *Symbiodinium minutum* has permanently condensed chromatin and its genome contains both eukaryotic histone genes and prokaryotic histone-like genes (Shinzato et al., 2014). Interestingly, the *S. minutum* genome lacks any chromatin-remodeling enzyme (Figure 4) suggesting that transcriptional regulation of its genes differs from known mechanisms. Of special interest are protozoan parasites, which provide insight into the evolution of transcription and chromatin dynamics. Many organisms belonging to these lineages (Microsporidia [*Edhazardia aedis*, *Encephalitozoon intestinalis*, and *Vavraia culicis*], Kinetoplastida [*Trypanosoma brucei*, *Trypanosoma cruzi*, and *Leishmania major*], Apicomplexa [*S. minutum*, *Perkinsus marinus*, *Cryptosporidium parvum*, *Plasmodium falciparum*, and *Toxoplasma gondii*] and *Giardia*) have a reduced set of chromatin remodelers (Figure 4), which may result from massive gene loss, commonly observed in parasites. The malaria parasite *P. falciparum* is a protist with a very AT-rich genome and with a disconnection between chromatin structure and gene expression (Westenberger et al., 2009). Intriguingly, some of the early branching parasitic protists like Kinetoplastida exert little control at the transcription level, which rather occurs post-transcriptionally (Kramer, 2012). Their protein-coding genes are arranged in long tandem arrays and transcribed as

long poly-cistrons (10–100 genes). Thus, these parasites have limited transcriptional regulation at the level of chromatin, and they also have different nucleosome arrangements and constitutions to support their complex lifestyle and to adapt to their environmental niche.

The histone variants H3.3, H2A.Z, and H2A.X are almost universally present indicating that they arose early in eukaryotic evolution (Malik and Henikoff, 2003). One would expect that histone variants and specific chromatin remodeling complexes acting upon these variants co-evolved in species. Support for this comes from budding yeast, which expresses only a single H3 protein resembling H3.3 and lacks ATRX (Figure 4). Certain histone lineages are categorized as outliers including: functionally specialized lineages, ancestral eukaryotic lineages that diverged early, and recent lineages subject to relaxed selection. Relaxed selective evolutionary constraints could account for the more rapid rate of histone evolution seen in Microsporidia (Malik and Henikoff, 2003). This strong divergence and accelerated evolution of histones might explain their limited set of chromatin remodeling enzymes (Figure 4). Clearly, detailed phylogenetic comparisons of chromatin remodelers and of their histone substrates provide testable hypotheses and further mechanistic insights.

Restructuring TBP-TATA and Liberating TBP

ATP-dependent remodeling is not unique to histone-DNA complexes. The inherently stable TBP-TATA complex is regulated directly by BTAf1/Mot1p that also uses ATP to mobilize the TBP at core promoters (Figure 2). BTAf1/Mot1p family members are also SWI2/SNF2-family ATPases and they bind to TBP in the presence or absence of DNA (Auble and Hahn, 1993; Auble et al., 1994; Timmers et al., 1992; Timmers and Sharp, 1991). BTAf1/Mot1p relaxes the DNA sequence-specificity of TBP to allow binding to non-TATA sequences (Gumbs et al., 2003; Klejman et al., 2005). BTAf1/Mot1p binds to TBP with its N-terminal HEAT/ARM repeats and contacts DNA upstream of TATA with its ATPase domain (Wollmann et al., 2011). BTAf1/Mot1p function is intimately linked to NC2. In living cells these factors together control the residence time of TBP on chromatin (de Graaf et al., 2010; Sprouse et al., 2008). The NC2 heterodimer consists of NC2 α and NC2 β , which interact via HFDs, resembling histones H2A and H2B (Kamada et al., 2001). NC2 inhibits PIC formation by competing with TFIIA and TFIIB for TBP binding (Goppelt et al., 1996; Meisterernst and Roeder, 1991; Mermelstein et al., 1996). Structural studies indicate that NC2 may embrace the TBP-TATA complex to close a ring around the DNA (Kamada et al., 2001). In vitro findings support that NC2 induces TBP sliding along the DNA (Schluesche et al., 2007).

Historically, BTAf1/Mot1p and NC2 have been studied in separation, but genome-wide mapping in yeast showed that binding profiles of Mot1p and NC2 strongly overlap (van Werven et al., 2008). Yeast strains with *ts*-mutations in *MOT1*, *NC2 α* , and *NC2 β* display similar alterations in mRNA expression (Dasgupta et al., 2002; Sikorski and Buratowski, 2009; Spedale et al., 2012; van Werven et al., 2008). Mot1p-TBP-NC2-TATA complexes can be disrupted in vitro as a result from Mot1p-mediated ATP hydrolysis (van Werven et al., 2008). Compared to TATA-less promoters, TBP turnover at TATA-containing promoters is relatively

high (van Werven et al., 2009). This is counterintuitive as the canonical TATA box represents DNA with the highest affinity for TBP (Hahn et al., 1989). We proposed that the bent TATA conformation induced by TBP binding could act as a “spring” for rapid BTAf1/Mot1p-NC2 mediated release from TATA boxes (Tora and Timmers, 2010). Auble and colleagues proposed models involving DNA-translocation by the ATPase moiety of BTAf1/Mot1p coupled to insertion of a latch from the HEAT repeat region into the concave surface of TBP to compete with DNA binding (Pereira et al., 2001; Viswanathan and Auble, 2011). The combined action of Mot1p and NC2 mobilizes TBP from intrinsically preferred TATA-containing promoters, which allows TBP redistribution to intrinsically disfavored TATA-less promoters (Zentner and Henikoff, 2013). This explains how Mot1p and NC2 repress SAGA-dependent TATA-containing genes and how they activate TFIID-dependent TATA-less genes (Spedale et al., 2012). It is interesting to note that a recent study on SAGA-bound TBP in yeast indicates that the concave surface of TBP remains largely accessible (Han et al., 2014), which may provide an entry zone for BTAf1/Mot1p and NC2. Given the strong sequence conservation between BTAf1/Mot1p, NC2, and TBP it seems likely that this is a common mechanism in eukaryotes.

TBP and Related Factors: Phylogenetics, Function, and Regulation

Proper TBP function is fundamental to the fidelity of transcriptional programs in both Archaea and Eukarya. The highly conserved C-terminal half of TBP consists of two symmetric pseudo-repeats (the TBP domain) folding into a saddle-shaped structure. While the convex surface interacts with proteins like TAFs, BTAf1/Mot1p, NC2, and basal transcription factors, the concave surface binds to the TATA box via the insertion of two pairs of phenylalanine to induce the bent conformation of TATA (Delgadillo et al., 2009). The evolutionary origin of the TBP domain can be traced back to the last universal common ancestor (LUCA) to Archaea and Eukarya and most likely resulted from an ancestral gene duplication and fusion event (Brinddefalk et al., 2013). TBP domains are present in proteins with diverse functions like DNA glycosylases and RNase III (Brinddefalk et al., 2013).

The compact nature of eukaryotic chromatin might have mandated a more stable DNA interaction of TBP compared to that in Archaea. Possibly, evolutionary acquirement of the critical phenylalanines provided stability to eukaryotic TBP-DNA complexes and resulted in the deformability of the TATA-sequences. Interestingly, promoter bending by eukaryotic and archaeal TBP and TFB/TFIIB occurs via molecularly distinct mechanisms (Gietl et al., 2014). The rapid on- and off-rates of archaeal TBP on DNA allows regulation directly at the recruitment stage. In line with this, archaeal transcription initiation is inhibited by sequence-specific regulators that compete with TBP and TFB for binding to the TATA box and BRE, or with RNAP for the site of transcription initiation (Reeve, 2003). Archaeal species living at high-temperature and/or high-salt concentrations increased the hydrophobicity of the TBP interior to withstand these extreme conditions (Koike et al., 2004).

Interestingly, most metazoan eukaryotes encode for multiple TBP paralogs, the TBP-related factors (TRFs) (Akhtar and

Veenstra, 2011; Levine et al., 2014; Müller et al., 2010). Independent duplication events gave rise to genes encoding insect-specific TRF1, metazoan-specific TRF2/TLF/TBPL1, and vertebrate-specific TBP2/TRF3/TBPL2 proteins. TRF1 associates with BRF in *Drosophila melanogaster* to form the TFIIB complex driving pol III-dependent transcription instead of TBP (Takada et al., 2000). The vertebrate-specific TBP2/TRF3 binds to the TATA box and interacts with TFIIA and TFIIB. TBP2/TRF3 can replace TBP for transcription in oocytes. During early development TBP levels increase and TBP2/TRF3 is actively degraded (Akhtar and Veenstra, 2011; Levine et al., 2014; Müller et al., 2010). TBP-like factor (TLF or TRF2) is the most distant paralog that evolved prior to the emergence of the bilateria and subsequent to the split between bilaterian and non-bilaterian animals (Duttke et al., 2014). TLF/TRF2 functions in male germ cell differentiation (male TLF/TRF2 null mice are sterile) and is essential for early embryogenesis in all non-mammalian metazoans studied thus far (Akhtar and Veenstra, 2011; Levine et al., 2014; Müller et al., 2010). TLF/TRF2 interacts with TFIIA and TFIIB, but lost the capacity to bind to the TATA box due to loss of two of the four phenylalanines required for TATA box recognition (Duttke et al., 2014; Teichmann et al., 1999). TLF/TRF2 is targeted to TATA-less promoters including the histone H1 promoter and it activates TCT- and DPE-containing promoters (Duttke et al., 2014; Isogai et al., 2007; Kedmi et al., 2014). The divergence in structure, expression, and function of TBP homologs explains their evolutionary retention. Thus far, most work has focused on TBP-containing complexes and the molecular mechanisms underlying the regulation of TBP paralogs remain to be elucidated.

Interestingly, some protists including *Giardia intestinalis*, *Cryptosporidium parvum*, *T. brucei*, *T. cruzi*, and *L. major* replaced multiple of the four critical phenylalanine residues in their single-copy TBP genes (Best et al., 2004; Das et al., 2005; Guillebault et al., 2002). Thus, these organisms must use different PIC assembly strategies, which still dependent on TBP but not on TATA box interactions. The promoter binding events are probably more dynamic, and to stabilize TBP-DNA interaction these organisms might depend more on the presence of other proteins, like TFIIA and TFIIB. However, this is not the case in *G. intestinalis*, because it seems to lack TFIIB (Best et al., 2004). Possibly, in this organism Brf1p, part of TFIIB, or a non-conserved protein with similar function, replaces TFIIB in pol II transcription. At present it is unclear how PIC assembly is achieved in these protozoan parasites and certain unicellular eukaryotes as they lack most of the basal transcription factors (Figure 1). Research in this area will be full of surprises.

BTA1/Mot1 and NC2: Phylogenetics, Function, and Regulation

TBP orthologs play crucial roles in all Archaea and Eukarya, but only eukaryotic genomes contain genes orthologous to BTA1/MOT1, NC2 α and NC2 β . Analogous to σ -factors in Bacteria, DNA sequence-specific regulators can compete with archaeal TBP for promoter binding. It is interesting to note that the TBP-interacting protein 26 (TIP26) from *Thermococcus kodakarensis* KOD1 can bind archaeal TBP inhibiting DNA binding (Yamamoto et al., 2006). Proteins with analogous functions to TIP26 might exist in other archaeal species. Alternatively, no additional fac-

tors could be required to disrupt archaeal TBP-DNA complexes as they are very dynamic intrinsically (Gietl et al., 2014).

While TBP regulation by BTA1/Mot1p and NC2 is well studied, their action toward the TBP paralogs of higher eukaryotes is not yet clear. Human BTA1 was found to interact with both *Caenorhabditis elegans* TRF2/TLF and *D. melanogaster* TRF1 (Pereira et al., 2001). In vitro transcription assays revealed that NC2 does not compete with TFIIA when bound to human TRF2 in contrast to TBP (Teichmann et al., 1999). This is an interesting area of study given the importance of TBP paralogs in germ cells and early embryogenesis (Akhtar and Veenstra, 2011; Duttke et al., 2014; Müller et al., 2010; Torres-Padilla and Tora, 2007).

We proposed previously that BTA1/Mot1p and NC2 act together on TBP, which implies co-occurrence of their genes across eukaryotes (van Werven et al., 2008). Indeed, testing this hypothesis revealed a clear co-occurrence and similar distribution of BTA1/MOT1, NC2 α , and NC2 β genes across the eukaryotic lineage (Figure 5). This provides strong evidence that these genes co-evolved. Interestingly, the Kinetoplastida and Apicomplexa protozoan parasites lack both BTA1/MOT1 and NC2 genes. Unfortunately, little is known about of transcriptional control in Apicomplexa. They contain a primitive transcription machinery lacking most of the TAFs and the basal transcription factors TFIIA and TFIIF (Meissner and Soldati, 2005). Typical eukaryotic promoter elements like TATA boxes are also absent. More is known about transcription regulation in Kinetoplastida. *G. intestinalis*, *T. brucei*, and *L. major* do not employ canonical TATA boxes for transcription initiation (Thomas et al., 2009). These species contain TBP homologs lacking the critical TATA-intercalating phenylalanine residues (Best et al., 2004; Guillebault et al., 2002; Ruan et al., 2004). Most likely, DNA interactions of these TBPs are weak and easily disrupted, which would obviate the need for TBP regulators BTA1/Mot1p and NC2. In *T. brucei* an alternative mechanism for TBP-promoter dissociation has been described, which involves PIC release from the promoter by TBP phosphorylation (Hope et al., 2014). We analyzed whether absence of the four phenylalanines in TBP is common in organisms lacking BTA1/MOT1 and NC2 orthologs. Organisms that lack the NC2 subunits, but have BTA1/Mot1p contain at least one TBP gene with all four intercalating phenylalanines (Tables S1, S2, S3, S4, S5, and S6). Interestingly, most of the organisms lacking BTA1/Mot1p do not contain a single TBP paralog with all four intercalating phenylalanines (Tables S1 and S2). In particular, the first phenylalanine (F193 in human TBP) is missing (6 out of 10). We propose that BTA1/Mot1p dependence is relaxed when TBP lacks the full complement of four phenylalanines. In contrast, organisms carrying the full set of genes orthologous to BTA1/MOT1, NC2 α , and NC2 β contain at least one TBP gene with all four intercalating phenylalanines. The only two exceptions to this rule (*B. natans* and *T. vaginalis*; Tables S5 and S6) carry an aromatic tyrosine, which could also intercalate into DNA. This persuasive correlation indicates co-evolution of stable TBP-DNA interactions with the enzymatic BTA1/Mot1p-NC2 machinery to enable dynamic transcriptional responses (Tora and Timmers, 2010; Viswanathan and Auble, 2011).

In summary, during evolution different strategies were developed to enable a dynamic binding of TBP orthologs and paralogs

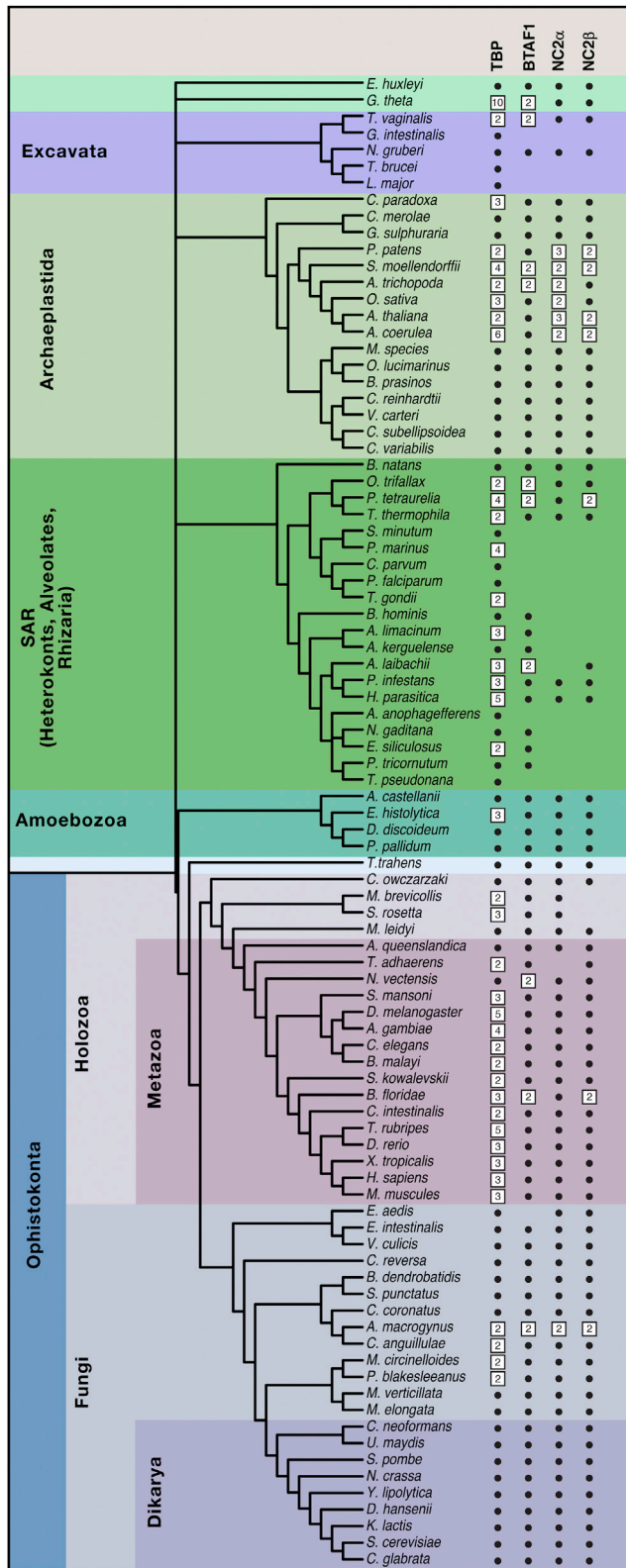


Figure 5. The Evolution of TBP and Its Direct Regulators

Schematic representation of the tree of life with a selection of eukaryotic species from the different supergroups indicated on the left. TATA-binding

protein (TBP) and its regulators are organized in different functional groups (TBP; NC2α; NC2β; BTA1) in a representation similar to Figure 4. These lists have been curated manually (see Supplemental Experimental Procedures).

Conclusions and Future Directions

The dynamic response of gene expression programs to cellular and environmental signals is a shared property of all living organisms. With the increase of genome size and complexity during the evolution of species came different mechanisms to ensure transcriptional dynamics and regulated accessibility of genomic sequences. In this review we discussed the function and evolutionary history of ATP-dependent enzymes controlling chromatin structure and PIC dynamics. Phylogenetic comparisons between Archaea and Eukarya reveal that histones and SWI2/SNF2 chromatin remodelers as well as TBP and BTA1/Mot1p originated from an ancestor common to both lineages.

During eukaryotic evolution remodelers diversified into four groups (SWI/SNF, ISWI/SNF2L, CHD/Mi-2, INO80), but not all eukaryotic genomes carry representatives of each group. Given their functional differences complete absence of a group (like ISWI/SNF2L in *S. pombe*) has direct consequences on chromatin structure (Pointner et al., 2012) and gene regulation pathways. ATP-dependent remodelers acquired additional (signature) domains for intra-molecular regulation and/or for chromatin interaction (Clapier and Cairns, 2009, 2012; Hargreaves and Crabtree, 2011). In almost all cases the enzymatic SWI2/SNF2 core has been decorated with many subunits, which modulate its activity, function, and/or localization. Cancer exome sequencing revealed that subunits of human SWI/SNF complexes are particularly prone to mutation and amplification in a variety of human cancers (Kadoch et al., 2013). From both fundamental and translational perspectives, it is important to determine evolutionary conservation and diversification of chromatin remodeler subunits. In addition, it would be interesting to analyze the evolutionary distribution of histone variants in relation to chromatin remodeling complexes.

Phylogenetic comparisons between the ATP-dependent BTA1/Mot1p and their TBP substrate reveal distinct patterns. Whereas all Eukarya contain one or more TBP genes, several species lack the BTA1/Mot1p gene. In most of these cases, no NC2 orthologs could be detected, which emphasizes the intimate link between BTA1/Mot1p and the NC2 complex in controlling TBP dynamics. Besides their TBP-regulatory domains BTA1/Mot1p and NC2α acquired additional domains during evolution, and their phylogenetic analysis may reveal accessory functions (Goppelt et al., 1996; Wollmann et al., 2011). It is striking to note that most organisms lacking BTA1/Mot1p express TBP orthologs, which are also lacking one or more of the four phenylalanines responsible for intercalating DNA.

Appreciating evolutionary relationships between chromatin and transcription proteins improves our overall understanding of gene and chromatin regulation principles. In these days, we are witnessing an ever-increasing wealth of genomic sequence data from present-day and extinct organisms, which offer unprecedented insight into evolutionary relationships between organisms and processes fundamental to life. Unfortunately, Theodosius Dobzhansky missed the birth of comparative genomics and of phylogenomics as he passed away 7 months after Sanger's first report on modern sequencing (Sanger et al., 1977) propelling this genomics revolution. Nevertheless, Dobzhansky realized the close association between environmental niche and the genome: "the environment presents challenges to living species, to which the latter respond by adaptive genetic changes" in (Dobzhansky, 1973). The fact, that the regulatory components of the transcription machinery are evolutionary malleable, should be no surprise as gene regulation steers many diverse processes as enzymatic adaption and organismal development. Understanding evolutionary conservation and diversity of these key components sheds light on the processes of adaptive gene expression and of organismal evolution itself.

Compared to the incredible airlift given by whole-genome sequencing in describing the genomic relatedness of organisms, description of their environmental niche remains grounded. For each organism, genome sequence and environment are inextricably linked, and we advocate attaching a standardized description of the environment to each genome sequence. These descriptions facilitate the linking of comparative zoology and phylogenetics to illuminate the fascinating 4.5 billion-year (bio-) chemical experiment underlying organismal evolution. We are sure that Dobzhansky would have been thrilled to partake in the current developments to understand the diversity of species.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures and six tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.04.033>.

ACKNOWLEDGMENTS

We are grateful to Geert Kops and members of the Timmers lab for discussion and insightful comments. We thank the anonymous reviewers for their astute comments. We thank Koen Dreijerink, Laszlo Tora, Finn Werner and Carl Wu for critical reading of the manuscript. The work is supported by the Netherlands Organization for Scientific Research (NWO) through ALW (grant #820.02.013 to HTMT).

REFERENCES

- Adelman, K., and Lis, J.T. (2012). Promoter-proximal pausing of RNA polymerase II: emerging roles in metazoans. *Nat. Rev. Genet.* 13, 720–731.
- Akhtar, W., and Veenstra, G.J. (2011). TBP-related factors: a paradigm of diversity in transcription initiation. *Cell Biosci.* 1, 23.
- Ammar, R., Torti, D., Tsui, K., Gebbia, M., Durbic, T., Bader, G.D., Giaever, G., and Nislow, C. (2012). Chromatin is an ancient innovation conserved between Archaea and Eukarya. *eLife* 1, e00078.
- Auble, D.T., and Hahn, S. (1993). An ATP-dependent inhibitor of TBP binding to DNA. *Genes Dev.* 7, 844–856.
- Auble, D.T., Hansen, K.E., Mueller, C.G., Lane, W.S., Thorner, J., and Hahn, S. (1994). Mot1, a global repressor of RNA polymerase II transcription, inhibits TBP binding to DNA by an ATP-dependent mechanism. *Genes Dev.* 8, 1920–1934.
- Bartholomew, B. (2014). Regulating the chromatin landscape: structural and mechanistic perspectives. *Annu. Rev. Biochem.* 83, 671–696.
- Basehoar, A.D., Zanton, S.J., and Pugh, B.F. (2004). Identification and distinct regulation of yeast TATA box-containing genes. *Cell* 116, 699–709.
- Becker, P.B., and Workman, J.L. (2013). Nucleosome remodeling and epigenetics. *Cold Spring Harb. Perspect. Biol.* 5, 5.
- Best, A.A., Morrison, H.G., McArthur, A.G., Sogin, M.L., and Olsen, G.J. (2004). Evolution of eukaryotic transcription: insights from the genome of *Giardia lamblia*. *Genome Res.* 14, 1537–1547.
- Blake, W.J., Balázs, G., Kohanski, M.A., Isaacs, F.J., Murphy, K.F., Kuang, Y., Cantor, C.R., Walt, D.R., and Collins, J.J. (2006). Phenotypic consequences of promoter-mediated transcriptional noise. *Mol. Cell* 24, 853–865.
- Brindfolk, B., Dessailly, B.H., Yeats, C., Orengo, C., Werner, F., and Poole, A.M. (2013). Evolutionary history of the TBP-domain superfamily. *Nucleic Acids Res.* 41, 2832–2845.
- Buratowski, S., Hahn, S., Guarente, L., and Sharp, P.A. (1989). Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* 56, 549–561.
- Chen, J., Zhang, Z., Li, L., Chen, B.C., Revyakin, A., Hajj, B., Legant, W., Dahan, M., Lionnet, T., Betzig, E., et al. (2014). Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell* 156, 1274–1285.
- Clapier, C.R., and Cairns, B.R. (2009). The biology of chromatin remodeling complexes. *Annu. Rev. Biochem.* 78, 273–304.
- Clapier, C.R., and Cairns, B.R. (2012). Regulation of ISWI involves inhibitory modules antagonized by nucleosomal epitopes. *Nature* 492, 280–284.
- Cler, E., Papai, G., Schultz, P., and Davidson, I. (2009). Recent advances in understanding the structure and function of general transcription factor TFIID. *Cell. Mol. Life Sci.* 66, 2123–2134.
- Côté, J., Quinn, J., Workman, J.L., and Peterson, C.L. (1994). Stimulation of GAL4 derivative binding to nucleosomal DNA by the yeast SWI/SNF complex. *Science* 265, 53–60.
- Darzacq, X., Shav-Tal, Y., de Turris, V., Brody, Y., Shenoy, S.M., Phair, R.D., and Singer, R.H. (2007). In vivo dynamics of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* 14, 796–806.
- Das, A., Zhang, Q., Palenchar, J.B., Chatterjee, B., Cross, G.A., and Bellofatto, V. (2005). Trypanosomal TBP functions with the multisubunit transcription factor tSNAP to direct spliced-leader RNA gene expression. *Mol. Cell. Biol.* 25, 7314–7322.
- Dasgupta, A., Darst, R.P., Martin, K.J., Afshari, C.A., and Auble, D.T. (2002). Mot1 activates and represses transcription by direct, ATPase-dependent mechanisms. *Proc. Natl. Acad. Sci. USA* 99, 2666–2671.
- Davis, J.L., Kunisawa, R., and Thorner, J. (1992). A presumptive helicase (MOT1 gene product) affects gene expression and is required for viability in the yeast *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 12, 1879–1892.
- de Graaf, P., Mousson, F., Geverts, B., Scheer, E., Tora, L., Houtsmuller, A.B., and Timmers, H.T. (2010). Chromatin interaction of TATA-binding protein is dynamically regulated in human cells. *J. Cell Sci.* 123, 2663–2671.
- Delgadillo, R.F., Whittington, J.E., Parkhurst, L.K., and Parkhurst, L.J. (2009). The TATA-binding protein core domain in solution variably bends TATA sequences via a three-step binding mechanism. *Biochemistry* 48, 1801–1809.
- Dinant, C., Luijsterburg, M.S., Höfer, T., von Bornstaedt, G., Vermeulen, W., Houtsmuller, A.B., and van Driel, R. (2009). Assembly of multiprotein complexes that control genome function. *J. Cell Biol.* 185, 21–26.
- Dion, M.F., Kaplan, T., Kim, M., Buratowski, S., Friedman, N., and Rando, O.J. (2007). Dynamics of replication-independent histone turnover in budding yeast. *Science* 315, 1405–1408.
- Dobzhansky, T. (1973). Nothing in biology makes sense except in the light of evolution. *Am. Biol. Teach.* 35, 126–129.

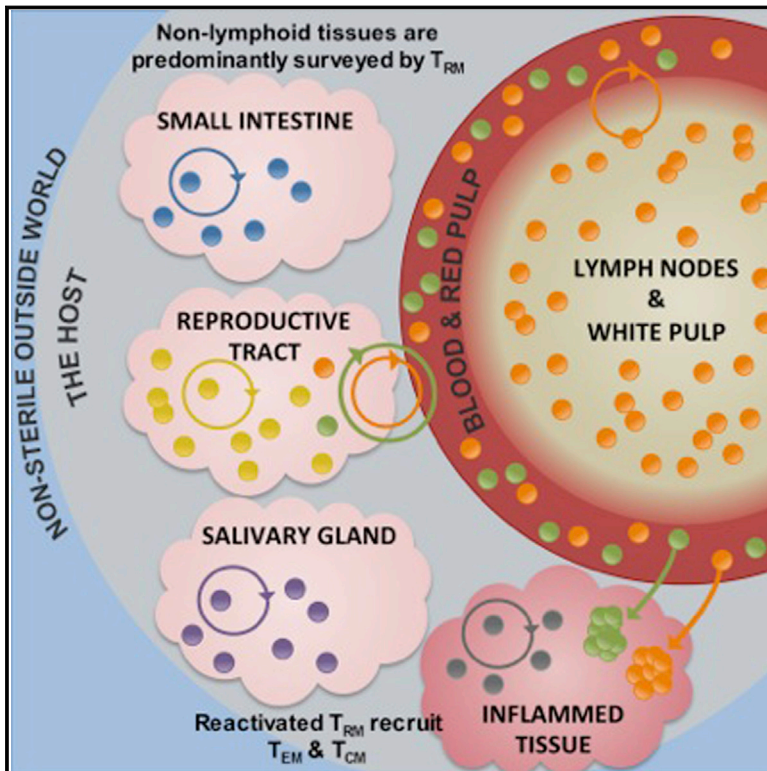
- Duttke, S.H., Doolittle, R.F., Wang, Y.L., and Kadonaga, J.T. (2014). TRF2 and the evolution of the bilateria. *Genes Dev.* 28, 2071–2076.
- Eisen, J.A., Sweder, K.S., and Hanawalt, P.C. (1995). Evolution of the SNF2 family of proteins: subfamilies with distinct sequences and functions. *Nucleic Acids Res.* 23, 2715–2723.
- Flaus, A., Martin, D.M., Barton, G.J., and Owen-Hughes, T. (2006). Identification of multiple distinct Snf2 subfamilies with conserved structural motifs. *Nucleic Acids Res.* 34, 2887–2905.
- Gazdag, E., Rajkovic, A., Torres-Padilla, M.E., and Tora, L. (2007). Analysis of TATA-binding protein 2 (TBP2) and TBP expression suggests different roles for the two proteins in regulation of gene expression during oogenesis and early mouse development. *Reproduction* 134, 51–62.
- Gentry, M., and Hennig, L. (2014). Remodelling chromatin to shape development of plants. *Exp. Cell Res.* 321, 40–46.
- Gietl, A., Holzmeister, P., Blombach, F., Schulz, S., von Voithenberg, L.V., Lamb, D.C., Werner, F., Tinnefeld, P., and Grohmann, D. (2014). Eukaryotic and archaeal TBP and TFB/TF(II)B follow different promoter DNA bending pathways. *Nucleic Acids Res.* 42, 6219–6231.
- Goppelt, A., Stelzer, G., Lottspeich, F., and Meisterernst, M. (1996). A mechanism for repression of class II gene transcription through specific binding of NC2 to TBP-promoter complexes via heterodimeric histone fold domains. *EMBO J.* 15, 3105–3116.
- Grohmann, D., and Werner, F. (2011). Recent advances in the understanding of archaeal transcription. *Curr. Opin. Microbiol.* 14, 328–334.
- Guillebault, D., Sasorith, S., Derelle, E., Wurtz, J.M., Lozano, J.C., Bingham, S., Tora, L., and Moreau, H. (2002). A new class of transcription initiation factors, intermediate between TATA box-binding proteins (TBPs) and TBP-like factors (TLFs), is present in the marine unicellular organism, the dinoflagellate *Cryptothecodinium cohnii*. *J. Biol. Chem.* 277, 40881–40886.
- Gumbs, O.H., Campbell, A.M., and Weil, P.A. (2003). High-affinity DNA binding by a Mot1p-TBP complex: implications for TAF-independent transcription. *EMBO J.* 22, 3131–3141.
- Hager, G.L., McNally, J.G., and Misteli, T. (2009). Transcription dynamics. *Mol. Cell* 35, 741–753.
- Hahn, S., Buratowski, S., Sharp, P.A., and Guarente, L. (1989). Yeast TATA-binding protein TFIID binds to TATA elements with both consensus and non-consensus DNA sequences. *Proc. Natl. Acad. Sci. USA* 86, 5718–5722.
- Han, Y., Luo, J., Ranish, J., and Hahn, S. (2014). Architecture of the *Saccharomyces cerevisiae* SAGA transcription coactivator complex. *EMBO J.* 33, 2534–2546.
- Hargreaves, D.C., and Crabtree, G.R. (2011). ATP-dependent chromatin remodeling: genetics, genomics and mechanisms. *Cell Res.* 21, 396–420.
- Hirschhorn, J.N., Brown, S.A., Clark, C.D., and Winston, F. (1992). Evidence that SNF2/SWI2 and SNF5 activate transcription in yeast by altering chromatin structure. *Genes Dev.* 6 (12A), 2288–2298.
- Hoopes, B.C., LeBlanc, J.F., and Hawley, D.K. (1998). Contributions of the TATA box sequence to rate-limiting steps in transcription initiation by RNA polymerase II. *J. Mol. Biol.* 277, 1015–1031.
- Hope, R., Ben-Mayor, E., Friedman, N., Voloshin, K., Biswas, D., Matas, D., Drori, Y., Günzl, A., and Michaeli, S. (2014). Phosphorylation of the TATA-binding protein activates the spliced leader silencing pathway in *Trypanosoma brucei*. *Sci. Signal.* 7, ra85.
- Isogai, Y., Keles, S., Prestel, M., Hochheimer, A., and Tjian, R. (2007). Transcription of histone gene cluster by differential core-promoter factors. *Genes Dev.* 21, 2936–2949.
- Iyer, L.M., Anantharaman, V., Wolf, M.Y., and Aravind, L. (2008). Comparative genomics of transcription factors and chromatin proteins in parasitic protists and other eukaryotes. *Int. J. Parasitol.* 38, 1–31.
- Jacob, F., and Monod, J. (1961). Genetic regulatory mechanisms in the synthesis of proteins. *J. Mol. Biol.* 3, 318–356.
- Jacobson, R.H., Ladurner, A.G., King, D.S., and Tjian, R. (2000). Structure and function of a human TAFII250 double bromodomain module. *Science* 288, 1422–1425.
- Johnson, T.A., Elbi, C., Parekh, B.S., Hager, G.L., and John, S. (2008). Chromatin remodeling complexes interact dynamically with a glucocorticoid receptor-regulated promoter. *Mol. Biol. Cell* 19, 3308–3322.
- Jun, S.H., Reichlen, M.J., Tajiri, M., and Murakami, K.S. (2011). Archaeal RNA polymerase and transcription regulation. *Crit. Rev. Biochem. Mol. Biol.* 46, 27–40.
- Juven-Gershon, T., and Kadonaga, J.T. (2010). Regulation of gene expression via the core promoter and the basal transcriptional machinery. *Dev. Biol.* 339, 225–229.
- Kadoch, C., Hargreaves, D.C., Hodges, C., Elias, L., Ho, L., Ranish, J., and Crabtree, G.R. (2013). Proteomic and bioinformatic analysis of mammalian SWI/SNF complexes identifies extensive roles in human malignancy. *Nat. Genet.* 45, 592–601.
- Kamada, K., Shu, F., Chen, H., Malik, S., Stelzer, G., Roeder, R.G., Meisterernst, M., and Burley, S.K. (2001). Crystal structure of negative cofactor 2 recognizing the TBP-DNA transcription complex. *Cell* 106, 71–81.
- Kedmi, A., Zehavi, Y., Glick, Y., Orenstein, Y., Ideses, D., Wachtel, C., Doniger, T., Waldman Ben-Asher, H., Muster, N., Thompson, J., et al. (2014). *Drosophila* TRF2 is a preferential core promoter regulator. *Genes Dev.* 28, 2163–2174.
- Kimura, H., and Cook, P.R. (2001). Kinetics of core histones in living human cells: little exchange of H3 and H4 and some rapid exchange of H2B. *J. Cell Biol.* 153, 1341–1353.
- Klejman, M.P., Zhao, X., van Schaik, F.M., Herr, W., and Timmers, H.T. (2005). Mutational analysis of BTAF1-TBP interaction: BTAF1 can rescue DNA-binding defective TBP mutants. *Nucleic Acids Res.* 33, 5426–5436.
- Koike, H., Kawashima-Ohya, Y., Yamasaki, T., Clowney, L., Katsuya, Y., and Suzuki, M. (2004). Origins of protein stability revealed by comparing crystal structures of TATA binding proteins. *Structure* 12, 157–168.
- Koonin, E.V. (2007). The Biological Big Bang model for the major transitions in evolution. *Biol. Direct* 2, 21.
- Koster, M.J., Yildirim, A.D., Weil, P.A., Holstege, F.C., and Timmers, H.T. (2014). Suppression of intragenic transcription requires the MOT1 and NC2 regulators of TATA-binding protein. *Nucleic Acids Res.* 42, 4220–4229.
- Kramer, S. (2012). Developmental regulation of gene expression in the absence of transcriptional control: the case of kinetoplastids. *Mol. Biochem. Parasitol.* 181, 61–72.
- Kwon, H., Imbalzano, A.N., Khavari, P.A., Kingston, R.E., and Green, M.R. (1994). Nucleosome disruption and enhancement of activator binding by a human SW1/SNF complex. *Nature* 370, 477–481.
- Lauberth, S.M., Nakayama, T., Wu, X., Ferris, A.L., Tang, Z., Hughes, S.H., and Roeder, R.G. (2013). H3K4me3 interactions with TAF3 regulate preinitiation complex assembly and selective gene activation. *Cell* 152, 1021–1036.
- Laurent, B.C., Yang, X., and Carlson, M. (1992). An essential *Saccharomyces cerevisiae* gene homologous to SNF2 encodes a helicase-related protein in a new family. *Mol. Cell. Biol.* 12, 1893–1902.
- Levine, M., Cattoglio, C., and Tjian, R. (2014). Looping back to leap forward: transcription enters a new era. *Cell* 157, 13–25.
- Lickwar, C.R., Mueller, F., Hanlon, S.E., McNally, J.G., and Lieb, J.D. (2012). Genome-wide protein-DNA binding dynamics suggest a molecular clutch for transcription factor function. *Nature* 484, 251–255.
- Lorch, Y., Maier-Davis, B., and Kornberg, R.D. (2014). Role of DNA sequence in chromatin remodeling and the formation of nucleosome-free regions. *Genes Dev.* 28, 2492–2497.
- Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389, 251–260.
- Malik, H.S., and Henikoff, S. (2003). Phylogenomics of the nucleosome. *Nat. Struct. Biol.* 10, 882–891.

- Maruyama, H., Harwood, J.C., Moore, K.M., Paszkiewicz, K., Durley, S.C., Fukushima, H., Atomi, H., Takeyasu, K., and Kent, N.A. (2013). An alternative beads-on-a-string chromatin architecture in *Thermococcus kodakarensis*. *EMBO Rep.* 14, 711–717.
- Meissner, M., and Soldati, D. (2005). The transcription machinery and the molecular toolbox to control gene expression in *Toxoplasma gondii* and other protozoan parasites. *Microbes Infect.* 7, 1376–1384.
- Meisterernst, M., and Roeder, R.G. (1991). Family of proteins that interact with TFIID and regulate promoter activity. *Cell* 67, 557–567.
- Meisterernst, M., Horikoshi, M., and Roeder, R.G. (1990). Recombinant yeast TFIID, a general transcription factor, mediates activation by the gene-specific factor USF in a chromatin assembly assay. *Proc. Natl. Acad. Sci. USA* 87, 9153–9157.
- Mermelstein, F., Yeung, K., Cao, J., Inostroza, J.A., Erdjument-Bromage, H., Egelson, K., Landsman, D., Levitt, P., Tempst, P., and Reinberg, D. (1996). Requirement of a corepressor for Dr1-mediated repression of transcription. *Genes Dev.* 10, 1033–1048.
- Moshkin, Y.M., Chalkley, G.E., Kan, T.W., Reddy, B.A., Ozgur, Z., van Ijcken, W.F., Dekkers, D.H., Demmers, J.A., Travers, A.A., and Verrijzer, C.P. (2012). Remodelers organize cellular chromatin by counteracting intrinsic histone-DNA sequence preferences in a class-specific manner. *Mol. Cell. Biol.* 32, 675–688.
- Müller, F., and Tora, L. (2014). Chromatin and DNA sequences in defining promoters for transcription initiation. *Biochim. Biophys. Acta* 1839, 118–128.
- Müller, F., Zaucker, A., and Tora, L. (2010). Developmental regulation of transcription initiation: more than just changing the actors. *Curr. Opin. Genet. Dev.* 20, 533–540.
- Munsky, B., Neuert, G., and van Oudenaarden, A. (2012). Using gene expression noise to understand gene regulation. *Science* 336, 183–187.
- Murawska, M., and Brehm, A. (2011). CHD chromatin remodelers and the transcription cycle. *Transcription* 2, 244–253.
- Narlikar, G.J., Sundaramoorthy, R., and Owen-Hughes, T. (2013). Mechanisms and functions of ATP-dependent chromatin-remodeling enzymes. *Cell* 154, 490–503.
- Neugeborn, L., and Carlson, M. (1984). Genes affecting the regulation of SUC2 gene expression by glucose repression in *Saccharomyces cerevisiae*. *Genetics* 108, 845–858.
- Newman, J.R., Ghaemmaghami, S., Ihmels, J., Breslow, D.K., Noble, M., DeRisi, J.L., and Weissman, J.S. (2006). Single-cell proteomic analysis of *S. cerevisiae* reveals the architecture of biological noise. *Nature* 441, 840–846.
- Papai, G., Weil, P.A., and Schultz, P. (2011). New insights into the function of transcription factor TFIID from recent structural studies. *Curr. Opin. Genet. Dev.* 21, 219–224.
- Pereira, L.A., van der Knaap, J.A., van den Boom, V., van den Heuvel, F.A., and Timmers, H.T. (2001). TAF(II)170 interacts with the concave surface of TATA-binding protein to inhibit its DNA binding activity. *Mol. Cell. Biol.* 21, 7523–7534.
- Pointner, J., Persson, J., Prasad, P., Norman-Axelsson, U., Strålfors, A., Khorosjutina, O., Krietenstein, N., Svensson, J.P., Ekwall, K., and Korber, P. (2012). CHD1 remodelers regulate nucleosome spacing in vitro and align nucleosomal arrays over gene coding regions in *S. pombe*. *EMBO J.* 31, 4388–4403.
- Poon, D., Campbell, A.M., Bai, Y., and Weil, P.A. (1994). Yeast Taf170 is encoded by MOT1 and exists in a TATA box-binding protein (TBP)-TBP-associated factor complex distinct from transcription factor IID. *J. Biol. Chem.* 269, 23135–23140.
- Ptashne, M. (2005). Regulation of transcription: from lambda to eukaryotes. *Trends Biochem. Sci.* 30, 275–279.
- Ranjan, A., Mizuguchi, G., FitzGerald, P.C., Wei, D., Wang, F., Huang, Y., Luk, E., Woodcock, C.L., and Wu, C. (2013). Nucleosome-free region dominates histone acetylation in targeting SWR1 to promoters for H2A.Z replacement. *Cell* 154, 1232–1245.
- Reeve, J.N. (2003). Archaeal chromatin and transcription. *Mol. Microbiol.* 48, 587–598.
- Rhee, H.S., and Pugh, B.F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* 483, 295–301.
- Ruan, J.P., Arhin, G.K., Ullu, E., and Tschudi, C. (2004). Functional characterization of a *Trypanosoma brucei* TATA-binding protein-related factor points to a universal regulator of transcription in trypanosomes. *Mol. Cell. Biol.* 24, 9610–9618.
- Sandelin, A., Carninci, P., Lenhard, B., Ponjavic, J., Hayashizaki, Y., and Hume, D.A. (2007). Mammalian RNA polymerase II core promoters: insights from genome-wide studies. *Nat. Rev. Genet.* 8, 424–436.
- Sandman, K., and Reeve, J.N. (2005). Archaeal chromatin proteins: different structures but common function? *Curr. Opin. Microbiol.* 8, 656–661.
- Sanger, F., Nicklen, S., and Coulson, A.R. (1977). DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 74, 5463–5467.
- Santangelo, T.J., Cubonová, L., James, C.L., and Reeve, J.N. (2007). TFB1 or TFB2 is sufficient for *Thermococcus kodakarensis* viability and for basal transcription in vitro. *J. Mol. Biol.* 367, 344–357.
- Schluesche, P., Stelzer, G., Piaia, E., Lamb, D.C., and Meisterernst, M. (2007). NC2 mobilizes TBP on core promoter TATA boxes. *Nat. Struct. Mol. Biol.* 14, 1196–1201.
- Shain, A.H., and Pollack, J.R. (2013). The spectrum of SWI/SNF mutations, ubiquitous in human cancers. *PLoS ONE* 8, e55119.
- Shinzato, C., Mungpakdee, S., Satoh, N., and Shoguchi, E. (2014). A genomic approach to coral-dinoflagellate symbiosis: studies of *Acropora digitifera* and *Symbiodinium minutum*. *Front Microbiol* 5, 336.
- Sikorski, T.W., and Buratowski, S. (2009). The basal initiation machinery: beyond the general transcription factors. *Curr. Opin. Cell Biol.* 21, 344–351.
- Song, F., Chen, P., Sun, D., Wang, M., Dong, L., Liang, D., Xu, R.M., Zhu, P., and Li, G. (2014). Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. *Science* 344, 376–380.
- Spedale, G., Meddens, C.A., Koster, M.J., Ko, C.W., van Hooff, S.R., Holstege, F.C., Timmers, H.T., and Pijnappel, W.W. (2012). Tight cooperation between Mot1p and NC2 β in regulating genome-wide transcription, repression of transcription following heat shock induction and genetic interaction with SAGA. *Nucleic Acids Res.* 40, 996–1008.
- Sprouse, R.O., Karpova, T.S., Mueller, F., Dasgupta, A., McNally, J.G., and Auble, D.T. (2008). Regulation of TATA-binding protein dynamics in living yeast cells. *Proc. Natl. Acad. Sci. USA* 105, 13304–13308.
- Stavreva, D.A., Müller, W.G., Hager, G.L., Smith, C.L., and McNally, J.G. (2004). Rapid glucocorticoid receptor exchange at a promoter is coupled to transcription and regulated by chaperones and proteasomes. *Mol. Cell. Biol.* 24, 2682–2697.
- Stern, M., Jensen, R., and Herskowitz, I. (1984). Five SWI genes are required for expression of the HO gene in yeast. *J. Mol. Biol.* 178, 853–868.
- Struhl, K. (1999). Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell* 98, 1–4.
- Takada, S., Lis, J.T., Zhou, S., and Tjian, R. (2000). A TRF1:BRF complex directs *Drosophila* RNA polymerase III transcription. *Cell* 101, 459–469.
- Teichmann, M., Wang, Z., Martinez, E., Tjernberg, A., Zhang, D., Vollmer, F., Chait, B.T., and Roeder, R.G. (1999). Human TATA-binding protein-related factor-2 (hTRF2) stably associates with hTFIIA in HeLa cells. *Proc. Natl. Acad. Sci. USA* 96, 13720–13725.
- Thomas, M.C., and Chiang, C.M. (2006). The general transcription machinery and general cofactors. *Crit. Rev. Biochem. Mol. Biol.* 41, 105–178.
- Thomas, S., Green, A., Sturm, N.R., Campbell, D.A., and Myler, P.J. (2009). Histone acetylations mark origins of polycistronic transcription in *Leishmania major*. *BMC Genomics* 10, 152.
- Timmers, H.T., and Sharp, P.A. (1991). The mammalian TFIID protein is present in two functionally distinct complexes. *Genes Dev.* 5, 1946–1956.
- Timmers, H.T., Meyers, R.E., and Sharp, P.A. (1992). Composition of transcription factor B-TFIID. *Proc. Natl. Acad. Sci. USA* 89, 8140–8144.
- Tirosh, I., and Barkai, N. (2008). Two strategies for gene regulation by promoter nucleosomes. *Genome Res.* 18, 1084–1091.

- Tora, L., and Timmers, H.T. (2010). The TATA box regulates TATA-binding protein (TBP) dynamics in vivo. *Trends Biochem. Sci.* 35, 309–314.
- Torres-Padilla, M.E., and Tora, L. (2007). TBP homologues in embryo transcription: who does what? *EMBO Rep.* 8, 1016–1018.
- van Werven, F.J., van Bakel, H., van Teeffelen, H.A., Altelaar, A.F., Koerkamp, M.G., Heck, A.J., Holstege, F.C., and Timmers, H.T. (2008). Cooperative action of NC2 and Mot1p to regulate TATA-binding protein function across the genome. *Genes Dev.* 22, 2359–2369.
- van Werven, F.J., van Teeffelen, H.A., Holstege, F.C., and Timmers, H.T. (2009). Distinct promoter dynamics of the basal transcription factor TBP across the yeast genome. *Nat. Struct. Mol. Biol.* 16, 1043–1048.
- Vannini, A., and Cramer, P. (2012). Conservation between the RNA polymerase I, II, and III transcription initiation machineries. *Mol. Cell* 45, 439–446.
- Vermeulen, M., Mulder, K.W., Denissov, S., Pijnappel, W.W., van Schaik, F.M., Varier, R.A., Baltissen, M.P., Stunnenberg, H.G., Mann, M., and Timmers, H.T. (2007). Selective anchoring of TFIID to nucleosomes by trimethylation of histone H3 lysine 4. *Cell* 131, 58–69.
- Viswanathan, R., and Auble, D.T. (2011). One small step for Mot1; one giant leap for other Swi2/Snf2 enzymes? *Biochim. Biophys. Acta* 1809, 488–496.
- Westenberger, S.J., Cui, L., Dharia, N., Winzeler, E., and Cui, L. (2009). Genome-wide nucleosome mapping of *Plasmodium falciparum* reveals histone-rich coding and histone-poor intergenic regions and chromatin remodeling of core and subtelomeric genes. *BMC Genomics* 10, 610.
- Williams, T.A., Foster, P.G., Cox, C.J., and Embley, T.M. (2013). An archaeal origin of eukaryotes supports only two primary domains of life. *Nature* 504, 231–236.
- Wollmann, P., Cui, S., Viswanathan, R., Berninghausen, O., Wells, M.N., Moldt, M., Witte, G., Butryn, A., Wendler, P., Beckmann, R., et al. (2011). Structure and mechanism of the Swi2/Snf2 remodeller Mot1 in complex with its substrate TBP. *Nature* 475, 403–407.
- Workman, J.L., and Roeder, R.G. (1987). Binding of transcription factor TFIID to the major late promoter during in vitro nucleosome assembly potentiates subsequent initiation by RNA polymerase II. *Cell* 51, 613–622.
- Yamaguchi, Y., Shibata, H., and Handa, H. (2013). Transcription elongation factors DSIF and NELF: promoter-proximal pausing and beyond. *Biochim. Biophys. Acta* 1829, 98–104.
- Yamamoto, T., Matsuda, T., Inoue, T., Matsumura, H., Morikawa, M., Kanaya, S., and Kai, Y. (2006). Crystal structure of TBP-interacting protein (Tk-TIP26) and implications for its inhibition mechanism of the interaction between TBP and TATA-DNA. *Protein Sci.* 15, 152–161.
- Yao, J., Munson, K.M., Webb, W.W., and Lis, J.T. (2006). Dynamics of heat shock factor association with native gene loci in living cells. *Nature* 442, 1050–1053.
- Zentner, G.E., and Henikoff, S. (2013). Mot1 redistributes TBP from TATA-containing to TATA-less promoters. *Mol. Cell. Biol.* 33, 4996–5004.

Quantifying Memory CD8 T Cells Reveals Regionalization of Immunosurveillance

Graphical Abstract



Authors

Elizabeth M. Steinert,
Jason M. Schenkel, ..., Peter J. Southern,
David Masopust

Correspondence

masopust@umn.edu

In Brief

Most anatomic compartments, including solid organs and vascular spaces, are patrolled by resident populations of memory CD8 T cells that stay local. The number of residents are underestimated by standard immunologic assays and indicate that immunosurveillance by CD8 T cells is highly regionalized.

Highlights

- Isolations underestimate memory CD8 T cells in nonlymphoid tissues (NLT)
- Together, memory CD8 T cells in NLTs and blood outnumber those in lymphoid organs
- Without inflammation, T_{CM} and T_{EM} surveillance of nonlymphoid tissues is minimal
- Each NLT is primarily surveyed by discrete resident populations that remain local



Quantifying Memory CD8 T Cells Reveals Regionalization of Immunosurveillance

Elizabeth M. Steinert,^{1,2,5} Jason M. Schenkel,^{1,2,5} Kathryn A. Fraser,^{1,2} Lalit K. Beura,^{1,2} Luke S. Manlove,^{2,3} Botond Z. Igyártó,^{2,4} Peter J. Southern,¹ and David Masopust^{1,2,*}

¹Department of Microbiology

²Center for Immunology

³Department of Laboratory Medicine and Pathology

⁴Department of Dermatology

University of Minnesota, Minneapolis, MN 55455, USA

⁵Co-first author

*Correspondence: masopust@umn.edu

<http://dx.doi.org/10.1016/j.cell.2015.03.031>

SUMMARY

Memory CD8 T cells protect against intracellular pathogens by scanning host cell surfaces; thus, infection detection rates depend on memory cell number and distribution. Population analyses rely on cell isolation from whole organs, and interpretation is predicated on presumptions of near complete cell recovery. Paradigmatically, memory is parsed into central, effector, and resident subsets, ostensibly defined by immunosurveillance patterns but in practice identified by phenotypic markers. Because isolation methods ultimately inform models of memory T cell differentiation, protection, and vaccine translation, we tested their validity via parabiosis and quantitative immunofluorescence microscopy of a mouse memory CD8 T cell population. We report three major findings: lymphocyte isolation fails to recover most cells and biases against certain subsets, residents greatly outnumber recirculating cells within non-lymphoid tissues, and memory subset homing to inflammation does not conform to previously hypothesized migration patterns. These results indicate that most host cells are surveyed for reinfection by segregated residents rather than by recirculating cells that migrate throughout the blood and body.

INTRODUCTION

A cardinal feature of the vertebrate adaptive immune system is the retention of a memory of past infections that enhances protective immunity in the event of reinfection. CD8 T cells are a principal component of this process and protect against those pathogens that invade intracellular compartments. Mechanistically, vertebrates maintain memory CD8 T cells that scan MHC I on the surface of host cells for the presence of pathogen-derived peptides. Recognition triggers infection

control. The efficiency achieved by this immunosurveillance depends upon the memory CD8 T cell population (1) magnitude relative to host cells and (2) location.

Quantification of the immune response is essential for our understanding of protective immunity and for evaluating vaccines. Limiting dilution assays suggested that pathogen-specific CD8 T cells were exceedingly rare among responding cells. However, technical innovations, such as the development of MHC I tetramers (Altman et al., 1996), revealed that antigen-specific CD8 T cell responses were 10- to 100-fold bigger than initially thought, precipitating a substantial revision in conceptualization of the immune response (Murali-Krishna et al., 1998).

Memory CD8 T cells are present within secondary lymphoid organs (SLO), blood, and the rest of the organism (nonlymphoid tissues [NLT], as well as primary lymphoid organs such as thymus and bone marrow). Landmark work, based on analysis of human blood, proposed that memory CD8 T cells could be parsed into two subsets based on their patterns of immunosurveillance. Central memory T cells (T_{CM}), defined by expression of lymph node homing molecules, putatively limit surveillance to SLO and are specialized for longevity and proliferation upon reinfection. Effector memory T cells (T_{EM}), defined by the absence of lymph node homing molecules, were thought to recirculate between blood, NLT, and lymph, thus surveying body surfaces and visceral organs that are often the initial portals of reinfection (Sallusto et al., 1999).

However, the T_{CM}/T_{EM} model failed to capture the true complexity of memory T cell diversity. It recently became clear that a third subset, termed tissue resident memory T cells (T_{RM}), resides in NLT without recirculating (Masopust and Schenkel, 2013; Mueller et al., 2013). Shortly after activation in SLO, this population seeds tissues, then differentiates in response to local environmental cues to adopt unique lineage-specific signatures (Casey et al., 2012; Mackay et al., 2013; Masopust et al., 2006). Importantly, the presence of T_{RM} at NLT sites of reinfection can accelerate pathogen elimination (Gebhardt et al., 2009; Jiang et al., 2012; Teijaro et al., 2011; Wu et al., 2014). Fundamentally, T_{RM} are defined by migration: they remain confined to one tissue without leaving and re-entering. Practically, cell migration patterns are laborious or impractical to define in animal models or humans, so phenotypic surface

markers have been substituted. The markers CD103 and CD69 are used to infer T_{RM} status, whereas the absence of both CD62L and CD69 expression defines NLT recirculating T_{EM} (Farber et al., 2014; Masopust and Schenkel, 2013). However, the fidelity of these markers has not been validated.

The emergence of T_{RM} has complicated the long standing paradigm of T cell-mediated immunosurveillance. It is no longer clear to what degree $CD8^+ T_{EM}$ recirculate through NLT and how immunological memories are apportioned between T_{RM} , T_{EM} , and T_{CM} , as each subset has not been quantified throughout the host. Previous identification of significant recirculation through major NLT (Klonowski et al., 2004) requires reassessment in light of recent discoveries of bloodborne populations contaminating even perfused tissues (Anderson et al., 2014). Moreover, while quantitative analyses typically depend on ex vivo isolation to determine memory CD8 T cell subset and phenotype, the accuracy of this approach has not been validated (Peaudecerf and Rocha, 2011; Selby et al., 1984). To address these gaps in the field, we performed a stringent and comprehensive quantitative analysis using migration properties to identify T_{RM} , T_{EM} , and T_{CM} populations. Our findings redress fundamental presumptions that inform models of immunosurveillance, T cell subsets, and protective immunity.

RESULTS

Isolations Underestimate Total Memory CD8 T Cells and Distort Distribution

Memory CD8 T cells are broadly distributed throughout the host organism, but the overall magnitude and anatomic apportionment of this population remain unclear and controversial (Ganusov and De Boer, 2007; Masopust et al., 2001; Peaudecerf and Rocha, 2011; Reinhardt et al., 2001; Rocha et al., 1991). To address this gap, we enumerated a single trackable memory CD8 T cell population established by a well-studied infection model in mice. To this end, we transferred naive lymphocytic choriomeningitis virus (LCMV)-specific Thy1.1⁺ P14 transgenic CD8 T cells into naive C57Bl/6J mice, which were then infected with LCMV (Armstrong strain). Animals were sacrificed 120–150 days later. These mice, referred to as P14 immune chimeras, were injected with α -CD8 α antibody (Ab) intravenously (i.v.) prior to sacrifice. The intravascular injection of α -CD8 α antibody was used in each experiment to distinguish i.v. Ab⁺ cells in vascular contiguous compartments (e.g., peripheral blood, spleen red pulp [RP], liver sinusoids, and lung capillaries) from i.v. Ab[−] CD8 T cells in the stroma and parenchyma of NLT and SLO (Anderson et al., 2014; Galkina et al., 2005). Cells were isolated from tissues by ex vivo dissociation (see Experimental Procedures) and then analyzed by flow cytometry.

Consistent with previous reports, we isolated ~6,000 P14 CD8 T cells from the female reproductive tract (FRT) (Nakanishi et al., 2009; Suvas et al., 2007). We also performed immunohistochemistry, taking advantage of the fact that the P14 LCMV system allows for identification of LCMV-specific cells in tissue sections via α -Thy1.1 Ab. Because ~240 7- μ m coronal sections could be acquired from the FRT, flow cytometry

data predicted ~25 P14 in a single section. But, we counted ~1,750 P14 per tissue section, suggesting discordance between flow cytometry and immunohistochemistry (data not shown).

For this reason, we developed an image-based quantitative immunofluorescence microscopy (QIM) strategy to compare the recovery of P14 memory CD8 T cells to what was actually present within the tissue (Figure 1A). For QIM, organ volumes of age-matched mice were determined by displacement. These values were consistent with available estimates from previous reports using a variety of methods (Doctor et al., 2010; Nutter et al., 1980; Scheller et al., 1994). Organs from P14 immune chimeras were also frozen, sectioned, and stained. Whole sections or large representative regions were imaged by immunofluorescence microscopy (see Experimental Procedures). Image size and section thickness were used to determine the portion of the whole organ represented in each image. This factor was used to extrapolate enumerations from large individual images to whole organs. Cell enumerations were then multiplied by 11/19 to correct for those cells that would be counted twice because they straddle two sections (Figures 1A and 1B). Importantly, the total number of nucleated cells in a given organ as determined by QIM was similar to that estimated by whole organ DNA content, assuming 6pg DNA per diploid cell (dos Anjos Pires et al., 2001), thus independently validating QIM accuracy (Figure 1B and Table 1).

QIM revealed that lymphocyte isolation from the FRT was inefficient, thus we tested whether isolation efficiencies varied among tissues by comparing these methods in many organs (Figure 1C). Many mucosal sites, including the stomach, lung, large intestine (LI) and FRT, contained 50- to 70-fold more α -CD8 α i.v. Ab[−] memory P14 CD8 T cells when evaluated by QIM as compared to cell isolation methods (Figure 1C and Table 1). While the density of memory P14 cells in skin was too low to evaluate (data not shown), QIM of other NLTs resulted in 6- to 27-fold higher estimates of P14s. Examination of SLOs, including the white pulp (WP) of the spleen and the mandibular lymph node (LN), resulted in the most efficient isolations with <2-fold differences observed between the two methods. These results demonstrate a wide discrepancy between cell isolation and QIM, suggesting that the most common method of enumeration (isolation) significantly underestimates the size of the memory CD8 T cell pool in NLT. Similar findings were observed when enumerating endogenous LCMV-specific memory CD8 T cells (without P14 transfers) in mice via in situ MHC I tetramer staining (Figure S1A) and also when analyzing CD8 β^+ T cells in human cervix (Figure S1B).

As memory CD8 T cells patrol and survey all nucleated cells for the presence of infection, we represented the total number of memory P14 CD8 T cells as determined by cell isolation (Figure 1D) or QIM (Figure 1E) per nucleated host cell (as determined by QIM) in LN, spleen, small intestine (SI), pancreas, stomach, FRT, and lung. Based on isolation methods, memory P14 CD8 T cells were calculated to be ~50- to 400-fold rarer in tissues than SLOs. QIM enumeration significantly altered this perceived immunosurveillance ratio and revealed that the density of sentinel memory CD8 T cells in NLT was within 8-fold of SLOs. This refinement in perspective could help explain how memory

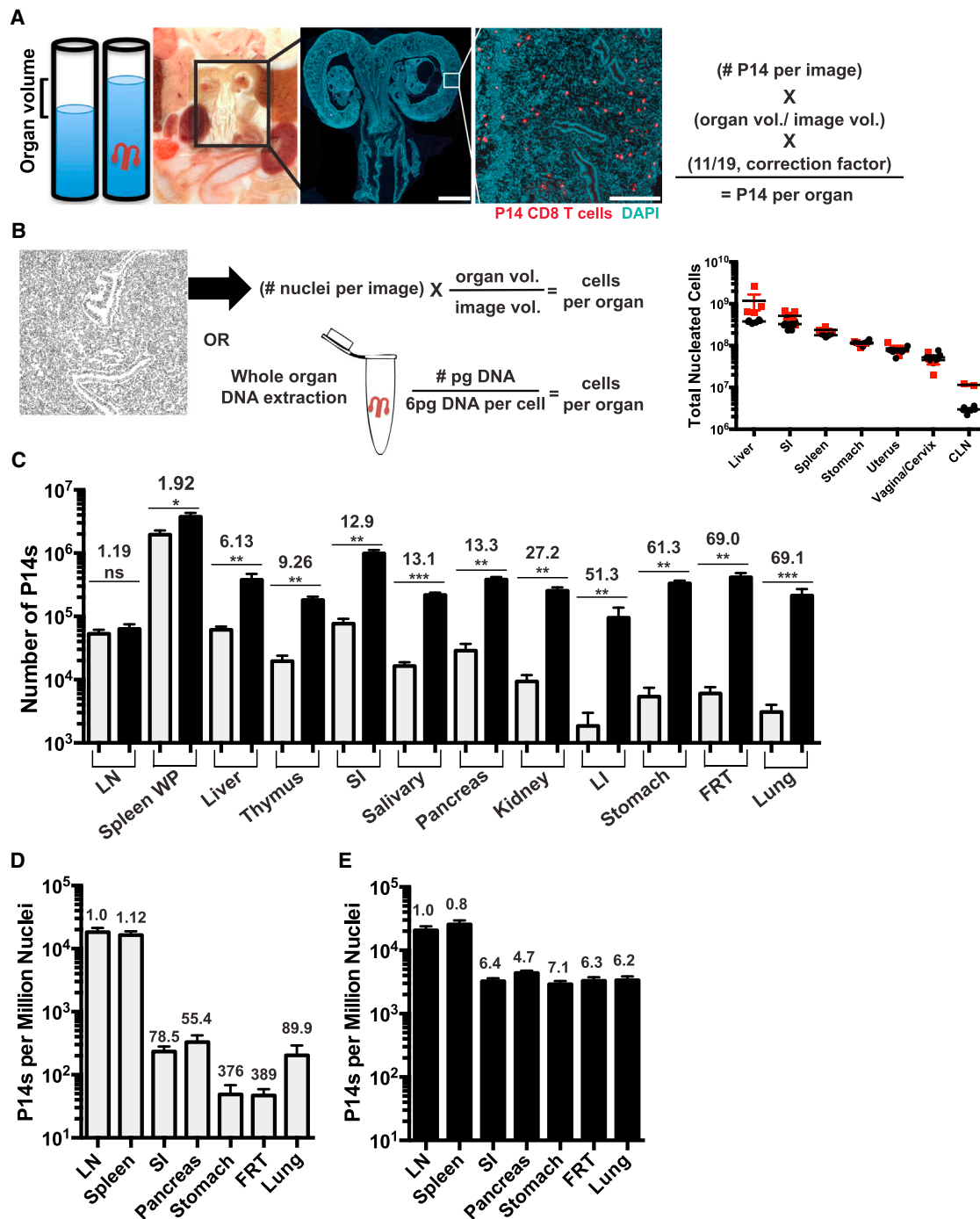


Figure 1. Isolations Underestimate Total Memory CD8 T Cells and Distort Distribution

(A and B) Quantitative Immunofluorescence Microscopy (QIM) methodology. (A) Organ volumes were determined by displacement. Tissue sections were stained for Thy1.1 (red) and DAPI (teal) to identify memory P14 CD8 T cells and nucleated cells 120–150 days after LCMV infection of C57Bl/6J mice. P14 counts per section were extrapolated to total organ volume and corrected to eliminate double counting. Whole FRT image scale bar, 2,000 μm ; cropped close up of FRT image scale bar, 250 μm . (B) Total DAPI⁺ nucleated cells by QIM were extrapolated to total organ volume (black circles) and validated independently by DNA extraction (red squares), $n = 4$.

(C) Comparison of α -CD8 α i.v. Ab⁺ P14 per tissue determined by cell isolation and flow cytometry (gray) or QIM (black).

(D and E) Total P14 frequency determined by (D) flow cytometry or (E) QIM relative to DAPI⁺ nucleated cells per organ as determined by QIM. Fold differences shown are relative to LN. $n \geq 6$, graphs show mean and SEM. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, Mann-Whitney-Wilcoxon test.

See also Figures S1 and S2.

Table 1. Enumeration of Memory P14 CD8 T Cells by Cell Isolation, Flow Cytometry, and QIM

Tissue	Fold Difference	Flow Cytometry		QIM			QIM and DNA
		P14 × 10 ⁴ ± SD × 10 ⁴	Average % of Total P14 i.v. Ab [−]	P14 × 10 ⁴ ± SD × 10 ⁴	Average % of Total P14 i.v. Ab [−]	P14 × 10 ³ ± SD × 10 ³ per 10 ⁶ Nuclei	Total Nucleated Cells × 10 ⁶ ± SD × 10 ⁶
	QIM/Flow						
Spleen			66.6		82.1	25.7 ^a ± 9.41	178 ^a ± 10.7 237 ± 30.2 ^b
White pulp (i.v. Ab [−])	1.92	195 ^a ± 79.7		375 ^a ± 142			
Red pulp (i.v. Ab ⁺)	0.84	93.2 ^a ± 32.7		78.5 ^a ± 20.5			
Mandibular lymph node	1.19	5.31 ^a ± 2.00	97.7	6.32 ^a ± 2.83	BD	20.7 ^a ± 7.74	2.98 ^a ± 0.48 11.5 ± 0.705 ^b
Thymus	9.26	1.96 ^a ± 1.11	96.6	18.1 ^a ± 5.61	99.7	2.86 ^a ± 0.32	77.37 ^a ± 25.7
Liver			14.8		16.9	5.42 ^a ± 1.17	378 ^a ± 29.9 1,180 ± 973 ^b
i.v. Ab [−]	6.13	6.12 ^a ± 2.05		37.5 ^a ± 22.0			
i.v. Ab ⁺	4.58	37.1 ^a ± 7.59		170 ^a ± 36.1			
Lung			9.06		25	3.36 ^a ± 1.27	282 ^a ± 45.2
i.v. Ab [−]	69.1	0.31 ^a ± 0.24		21.3 ^a ± 15.3			
i.v. Ab ⁺	13.8	5.43 ^a ± 6.42		75.2 ^a ± 53.7			
Kidney			46.3		83.1	1.99 ^a ± 0.697	157 ^a ± 25.0
i.v. Ab [−]	27.2	0.945 ^a ± 0.64		25.4 ^a ± 8.27			
i.v. Ab ⁺	5.02	1.03 ^a ± 0.76		5.17 ^a ± 1.95			
Pancreas	13.3	2.87 ^a ± 2.11	94.8	37.9 ^a ± 9.19	99.7	4.37 ^a ± 0.95	86.9 ^a ± 11.5
Salivary gland							
Serous	13.1	1.65 ^a ± 0.59	99.8	21.6 ^a ± 5.61	BD	10.3 ^a ± 2.64	21.2 ^a ± 2.58
Mucous	NA	NA	NA	7.81 ^a ± 2.35	BD	5.43 ^a ± 1.41	14.4 ^a ± 1.42
FRT	69.0	0.603 ^a ± 0.41	90.2				
Uterus				25.6 ^a ± 4.61	BD	3.51 ^a ± 0.958	75.7 ± 15.8 85.4 ± 26.2 ^b
Cervix/ vagina				16.0 ^a ± 8.39	BD	2.95 ^a ± 1.02	52.4 ^a ± 14.2 45.1 ± 19.8 ^b
SI						3.24 ^a ± 0.91	328 ^a ± 92.9 517 ± 176 ^b
IEL	6.10	3.81 ^a ± 2.02	99.8	23.2 ^a ± 12.7	BD		
LP + muscle ^c	18.6	4.06 ^a ± 1.70	99.1				
LP				74.3 ^a ± 22.5	BD		
Muscle				1.04 ^a ± 1.05	BD		
LI						0.81 ^a ± 0.41	122 ^a ± 12.3
IEL	41.3	0.034 ^a ± 0.018	84.10	1.39 ^a ± 0.75	BD		
LP + ILF ^c	68.1	0.12 ^a ± 0.073	82.46				
LP				7.59 ^a ± 3.65	BD		
ILF				0.51 ^a ± 0.63	BD		
Stomach						2.91 ^a ± 0.92	118 ^a ± 14.9 113 ± 17.4 ^b
IEL	17.5	0.35 ^a ± 0.45	91.9	6.17 ^a ± 2.27	BD		

(Continued on next page)

Table 1. Continued

Tissue	Fold Difference	Flow Cytometry		QIM		QIM and DNA	
		P14 × 10 ⁴ ± SD × 10 ⁴	Average % of Total P14 i.v. Ab ⁻	P14 × 10 ⁴ ± SD × 10 ⁴	Average % of Total P14 i.v. Ab ⁻	P14 × 10 ³ ± SD × 10 ³ per 10 ⁶ Nuclei	Total Nucleated Cells × 10 ⁶ ± SD × 10 ⁶
LP + SM + ME ^c	122	0.22 ^a ± 0.157	95.1				
LP				20.0 ^a ± 7.22	BD		
SM				3.05 ^a ± 0.92	BD		
ME				3.91 ^a ± 1.12	BD		
Peripheral blood	NA	13.4 ^a ± 6.33	100	NA	NA	NA	NA

Naive 5 × 10⁴ Thy1.1+ P14 CD8 T cells were transferred to C57Bl/6J mice, which were infected 1 day later with 2 × 10⁵ pfu LCMV Armstrong i.p. Approximately 120–150 days later, 3 min prior to sacrifice, mice were injected i.v. with α-CD8α antibody to discriminate the blood and marginated pool (i.v. Ab⁺) from parenchymal P14 (i.v. Ab⁻).

BD, below detection; NA, not available; LP, lamina propria; IEL, intraepithelial lymphocytes; SM, submucosa; ME, muscularis externa; ILF, isolated lymphoid follicle; FRT, whole female reproductive tract.

^aIndicates the average number of P14 or total nucleated cells per tissue derived from cell isolation and flow cytometry or QIM. Kidney accounts for both kidneys, salivary gland reports for both lobes, uterus includes both uterine horns, and mandibular lymph node enumerates a single unpaired lymph node. Peripheral blood enumeration is extrapolated to 1.74 ml of blood, based on average body weight of mice used in this study. Data from six or more mice.

^bIndicates number of nucleated cells (±SD) as determined by DNA extraction.

^cIndicates compartments indistinguishable by digestions and flow cytometry.

CD8 T cells within NLT can be sufficiently abundant to be first responders against anamnestic infections (Masopust and Schenkel, 2013; Mueller et al., 2013).

Isolation Efficiency Is Biased by Tissue Compartment and Cell Phenotype

Because cell isolation methods failed to capture most cells from NLT, we asked whether isolation efficiency varied among memory CD8 T cells with different phenotypes or between compartments within organs, thus further distorting the representation of the memory CD8 T cell population composition and location. Using intravascular α-CD8α Ab, we found that the blood and marginated pool (BMP) of lymphocytes (i.v. Ab⁺) within kidney and lung were more readily isolated than those within the tissue (i.v. Ab⁻) (Figures 2A and 2B). This was also true of splenic RP (i.v. Ab⁺) compared to splenic WP (i.v. Ab⁻) (Table 1).

We next investigated if lymphocyte extraction efficiency differed between histologically distinct mucosal compartments. To this end, we separated analyses of memory CD8 T cells isolated or imaged from stomach and SI into fractions localized above the basement membrane (intraepithelial lymphocytes [IEL]) or cells contained within the collagen matrix subjacent to the epithelium (lamina propria [LP] lymphocytes) (Figure 2C). As shown in Figure 2D and Table 1, P14 memory CD8 T cells are more efficiently recovered from epithelium than the lamina propria.

We next examined whether lymphocyte isolation misrepresented the proportion of mucosal memory CD8 T cell subsets as defined by phenotype. We focused on the FRT because it contains both CD103⁺ and CD103⁻ memory P14 CD8 T cells (Figure 2E), and CD103 is one marker used to define T_{RM}.

As shown in Figure 2F, cell isolation from the FRT over-represents the proportion of P14 memory CD8 T cells that express CD103. This bias may also have an anatomic basis (as in Figure 2D) as CD103⁺ cells are enriched within epithelium relative to lamina propria (Figure 2G). Taken together, these results indicate that lymphocyte isolation from NLT misrepresents memory CD8 T cell distributions by location and phenotype.

Most Memory CD8 T Cells in NLT Are T_{RM}

A broad and accurate accounting of the anatomic distribution of a memory CD8 T cell population, delineated into resident (T_{RM}) versus recirculating (T_{EM} and T_{CM}) subsets, has not previously been performed. Moreover, since the identification of T_{RM} as a distinct lineage (previously T_{RM} were conflated with recirculating T_{EM}), it remains unclear what contribution each population makes to the overall NLT memory T cell pool and how these populations compare numerically with memory T cells positioned within SLOs. We first interrogated this issue by quantifying the proportion of memory CD8 T cells that were resident after LCMV infection. The vasculature of P14 immune chimeras (90 days after infection, generated as in Figure 1) was conjoined to that of naive mice via parabiosis surgery. Thirty days later, we tested whether memory P14 CD8 T cells equilibrated between immune and naive parabiont organs, or whether disequilibrium was maintained which indicates residence (Figure 3A). As preliminary evidence indicated that flow cytometry preferentially underestimated T_{RM} as compared to recirculating T_{EM} (data not shown), we utilized the more precise QIM approach for this analysis.

Initially, we restricted analysis to P14 memory CD8 T cells that were not permissive to i.v. Ab staining. SLOs maintained very

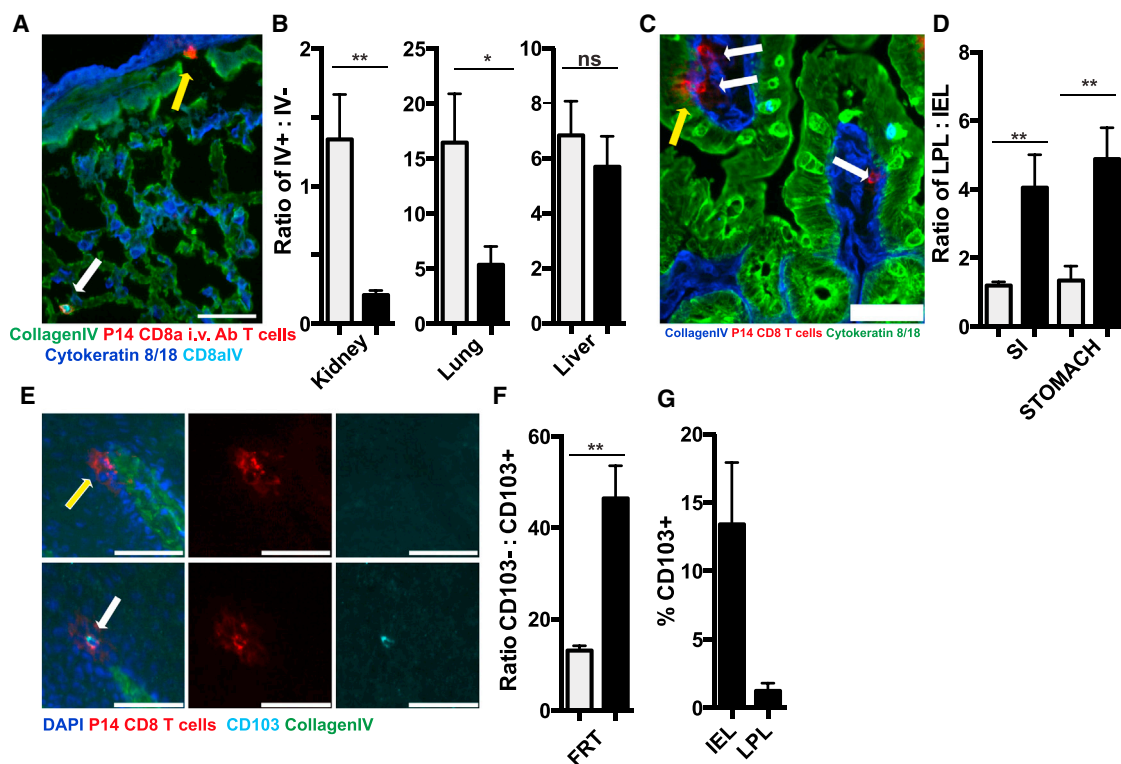


Figure 2. Isolation Efficiency Is Biased by Tissue Compartment and Cell Phenotype

P14 immune chimeras were analyzed 120–150 days after LCMV infection.

(A) Representative image of CD8 α i.v. Ab⁺ (white arrow) or CD8 α i.v. Ab⁻ (yellow arrow) P14 CD8 T cells in lung. CD8 α i.v. Ab (teal), Thy1.1⁺ P14 (red), Collagen IV (green), and Cytokeratin 8/18 (Blue). Scale bar, 50 μ m.

(B) Ratio of i.v. Ab⁺ to i.v. Ab⁻ P14s by flow cytometry (gray) and QIM (black) methodology.

(C) Representative image of P14 CD8 T cell in small intestine epithelium (intraepithelial lymphocyte [IEL] indicated by yellow arrow) and lamina propria (lamina propria lymphocyte [LPL] indicated by white arrows). Thy1.1⁺ P14 (red), Collagen IV (blue), and Cytokeratin 8/18 (Green). Scale bar, 50 μ m.

(D) Ratio of LPL to IEL P14 by flow cytometry (gray) and QIM (black).

(E) Representative image of CD103⁻ (top panels) and CD103⁺ (bottom panels) P14 CD8 T cells in vaginal epithelium. CD103 (teal), Thy1.1⁺ P14 (red), Collagen IV (green), and DAPI (blue). Scale bar, 50 μ m.

(F) Ratio of CD103⁻ to CD103⁺ P14s by flow cytometry (gray) and QIM (black) in FRT.

(G) Percent of vaginal IEL or LPL P14 expressing CD013, determined by QIM. n \geq 6, graphs show mean and SEM. *p < 0.05, **p < 0.01, ***p < 0.001, Mann-Whitney-Wilcoxon test.

See also Figure S2.

little disequilibrium between immune and naive parabionts, consistent with the previous observation that they contain only small fractions of T_{RM} after LCMV infection (Schenkel et al., 2014). In contrast, the vast majority of memory P14 CD8 T cells within almost all NLT examined were T_{RM}, as they exhibited little to no evidence of infiltration into the NLTs of naive parabionts (Figures 3B and 3C). Indeed, liver was the only NLT that supported substantive levels of memory CD8 T cell migration, although even in this case, ~55% of i.v. Ab⁻ P14 CD8 T cells were resident.

The distribution of T lymphocytes and particular memory subsets remains uncertain and debated, in part due to technical issues of quantifying cell numbers in tissues, identification of antigen-specific populations with a known history of stimulation, and bona fide analyses of cell recirculation. As QIM, parabiosis, and our focus on a single but identifiable population (P14, 120 days after LCMV infection in mice) overcome these hurdles,

we summated the parabiosis data from each NLT, revealing that the vast majority of nonlymphoid memory P14 are in fact T_{RM}, not recirculating T_{EM} (Figure 3D). Further, we then leveraged these approaches to generate a global representation of the apportionment of a memory CD8 T cell population throughout the visceral compartments of the organism. These data, shown in Figure 3E, support several conclusions. Less than half of the memory P14 pool was localized to SLO, spleen WP and LN (extrapolating mandibular LN data to the 37 macroscopic LNs in mice) (Van den Broeck et al., 2006). This was due to the fact that NLT contained more cells than expected based on previous cell isolation-dependent methods and also because of the surprising abundance of memory P14 contained within the BMP, a compartment that has not been enumerated in previous studies. Indeed, peripheral blood (from which many estimates of total blood lymphocytes are extrapolated) actually contained <4% of the memory P14 within the total bloodborne

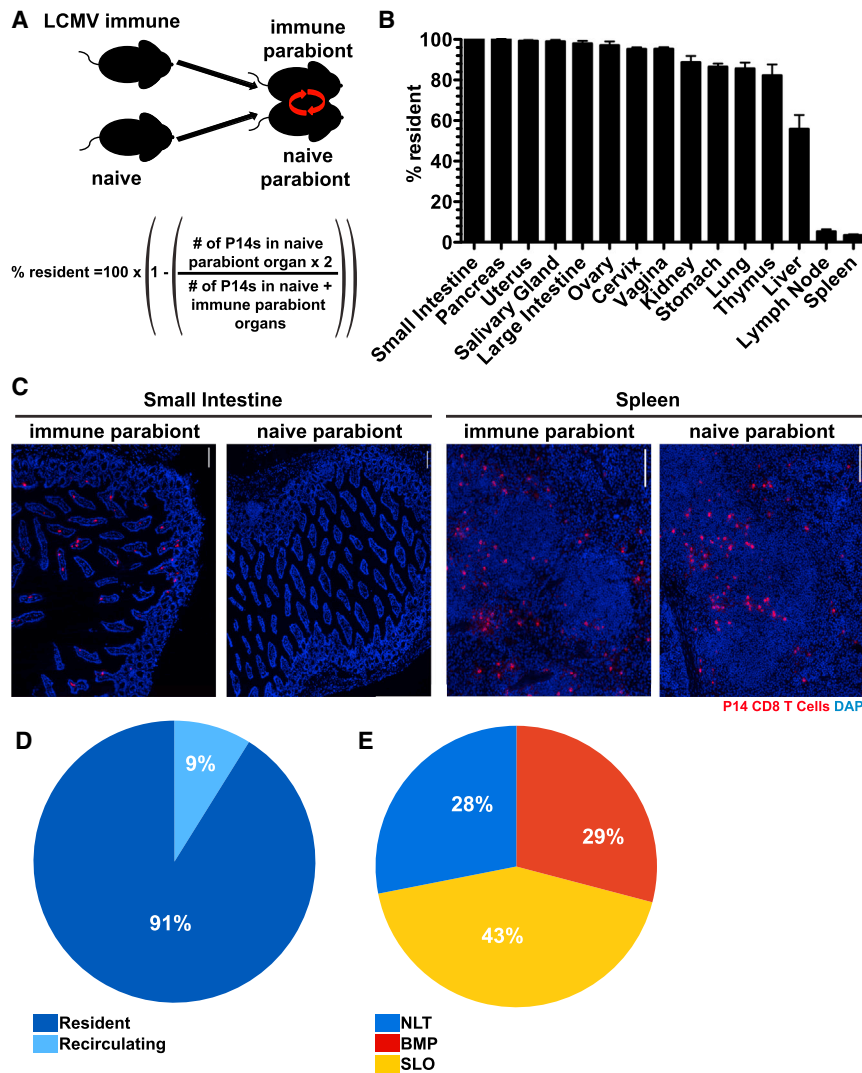


Figure 3. The Majority of Memory CD8 T Cells in NLT Are T_{RM}

(A) Ninety days after infection with LCMV Armstrong, P14 immune chimeras were conjoined to naive C57BL/6 mice using parabiosis.

(B–E) P14 immune chimeras conjoined to naive C57BL/6 mice were analyzed 30 days after parabiosis surgery. (B) Thirty days after parabiosis surgery the fraction of resident memory P14 CD8 T cells was calculated for the indicated tissues. $n = 3$, representative of nine mouse pairs from three independent experiments. Graphs show mean and SEM. (C) Representative images of P14 CD8 T cells in the small intestines and spleens of LCMV immune and naive parabionts, P14s (red) and DAPI (blue). Scale bar, 50 μm . (D) Distribution of resident and recirculating P14 CD8 T cells in nonlymphoid organs calculated by QIM. (E) P14 immune chimeras were analyzed 120–150 days after LCMV infection to determine the distribution of P14 CD8 T cells in secondary lymphoid organs (SLO), nonlymphoid tissues (NLT, including i.v. Ab^- cells within liver, lung, kidney, pancreas, salivary gland, uterus, vagina and cervix, small intestine, large intestine, stomach, and thymus) and circulating blood and marginated pool (BMP) (includes i.v. Ab^+ cells from all tissues examined), $n \geq 6$. Cell numbers from all tissues were calculated by QIM, except circulating blood, which was enumerated by cell isolation and flow cytometry. See also Figure S2.

population, particularly due to the magnitude or increased density of lymphocytes within spleen red pulp, lung and liver vasculature (Table 1). These data provide the most extensive quantitative characterization of a single memory CD8 T cell population to date and revise perceptions of migration and distribution.

Memory CD8 T Cell Migration Is Compartmentally Restricted within NLT

We next used the advantages of imaging analyses to test whether memory CD8 T cell entry during the memory phase of the response was selective for certain tissues within nonlymphoid organs. As shown in Figure 4A, mucosal organs could be segregated into three patterns of memory P14 migration, those in which there was: (1) no migration to mucosal epithelia or LP, (2) no migration to mucosal epithelia but limited migration to LP, submucosa, and muscularis externa, and (3) limited migration to both epithelia and LP. In the thymus, the medulla, but not cortex, was permissive to memory CD8 T cell recirculation.

These results suggested that memory CD8 T cell migration differs between compartments within nonlymphoid organs, although T_{RM} dominate all compartments. We next focused our analyses on the i.v. Ab^+ BMP in liver and kidney, which includes cells within sinusoids and glomeruli (Anderson et al., 2014). We observed that 35%–60% of the marginated pool was T_{RM} even within the vascular compartments of these organs (Figure 4D). These data indicate that migration properties vary by compartment within NLT and that T_{RM} are not exclusively localized to the parenchyma of tissues.

CD69 Is an Imperfect Marker of Tissue Residence

Given the impracticality of performing bona fide migration studies, the C-type lectin CD69 has become the defining marker for distinguishing T_{RM} from recirculating T_{EM} because it antagonizes the sphingosine 1-phosphate receptor 1 (S1PR1) that promotes egress via lymphatics and is necessary for T_{RM} maintenance in epidermis (Farber et al., 2014; Mackay et al., 2013). We tested whether CD69 expression was stringently predictive of recirculation properties. Only 25%–75% of the memory P14 cells in pancreas, salivary gland (SG), and FRT expressed CD69 (Figure 5A) even though almost all cells from these organs were T_{RM} (Figure 3B). This demonstrates that CD69[−] cells can also be functionally resident, a result that extends to the vascular compartments of the kidney and liver (Figures 5B–5D). Thus,

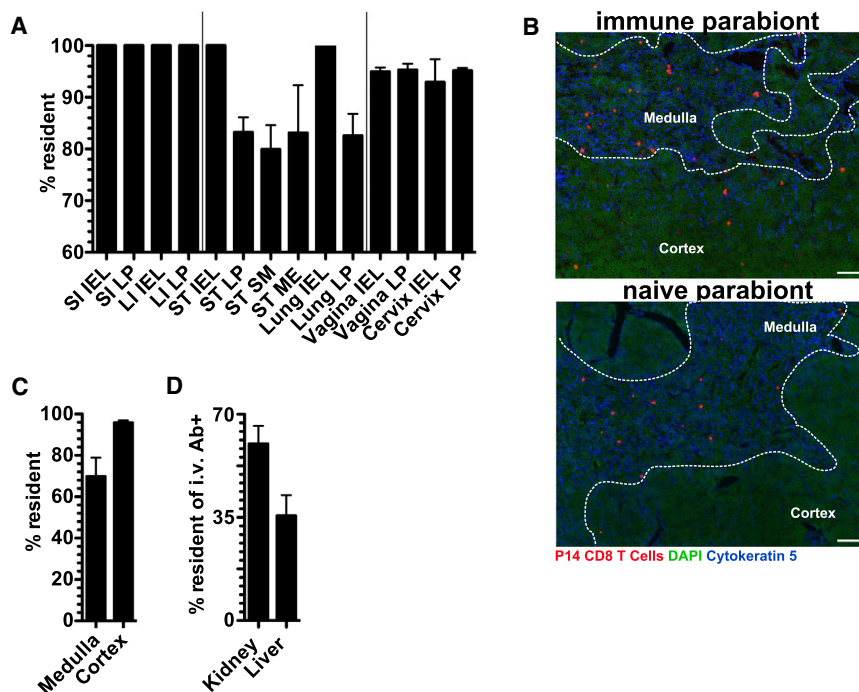


Figure 4. Memory CD8 T Cell Migration Is Compartmentally Restricted within NLT

P14 immune chimeras conjoined to naive C57BL/6 mice (as in Figure 3) were analyzed 30 days after parabiosis surgery.

(A) The fraction of P14 CD8 T cells that are resident in the indicated tissue compartments, small intestine (SI), large intestine (LI), stomach (ST), epithelium (IEL), lamina propria (LP), submucosa (SM), and muscularis externa (ME).

(B) Representative thymus images in immune and naive parabionts. P14 CD8 T cells (red), DAPI (green), and Cytokeratin 5 (blue). Scale bar, 50 μ m.

(C) Percent of P14 CD8 T cells that are resident in the thymus, medulla, and cortex.

(D) Percent of i.v. Ab⁺ P14 CD8 T cells that are resident within the kidney and liver. $n = 3$, representative of nine mouse pairs from three independent experiments. Graphs show mean and SEM. See also Figure S2.

CD69 is not a definitive marker to distinguish recirculating cells from T_{RM} .

CD69 is known to be induced on T_{RM} precursors upon migration into tissues during the effector phase of immune responses, putatively by tissue-derived instructional cues (Casey et al., 2012; Lee et al., 2011; Masopust et al., 2006). However, we observed CD69 expression among T_{RM} within the BMP of the liver and kidney, suggesting that parenchymal localization is not a requirement. Indeed, we even detected CD69⁺ memory P14 CD8 T cells within the large bore vessels of the liver of immune (but not naive) parabionts (Figure 5E). Taken together, in the steady state most CD69⁺ memory CD8 T cells are T_{RM} , but many T_{RM} are not CD69⁺.

Migration of Memory CD8 T Cell Subsets

Evidence for equilibration of memory CD8 T cells in non-lymphoid tissues fails to discriminate between bona fide recirculating T_{EM} versus the possibility that a few T_{EM} or T_{CM} continue to seed NLT and form T_{RM} long after immunization (i.e., a one-way trip). Because leukocytes use lymphatics to exit tissues, we examined whether we could observe evidence of memory P14 CD8 T cells within lymphatic vessels (visualized by Lyve-1 staining) of naive parabionts. We focused on FRT and SG due to the prominent nature of the lymphatic collecting ducts in these organs (Figures 6A–6C). Figure 6C of a representative FRT image shows that P14 memory CD8 T cells could indeed be visualized within lymphatic vessels. In each mouse, we visualized ~ 100 lymphatic vessel-bound P14 CD8 T cells in both FRT and SG when three to four sections were combined for analysis.

Quantitative analysis indicated that $\sim 20\%$ of P14 CD8 T cells that entered SG and FRT of naive parabionts during the mem-

ory phase of the response could be localized to lymphatic vessels (Figures 6A and 6B). These data provide strong evidence that a substantive fraction of P14 CD8 T cells that entered these NLT tissues

during the memory phase of the immune response were bona fide T_{EM} that exited these tissues after entry (even though T_{RM} represented the dominant fraction of the overall memory CD8 T cell population in these tissues, see Figures 3 and 4). Phenotypic analysis indicated that memory P14 CD8 T cells in lymphatic vessels were exclusively CD69[−] (Figures 6A and 6B). While this has not previously been reported, we were able to detect a population of CD69⁺ P14 CD8 T cells that had migrated to the FRT and SG of naive parabionts during the memory phase of the immune response, 90–120 days after infection.

Paradigmatically, T_{EM} recirculate through NLT or respond to NLT sites of inflammation, while T_{CM} limit recirculation to SLO (Sallusto et al., 1999). However, this hypothesis has not been rigorously tested. Parabiosis allowed us to identify bona fide CD69[−] memory CD8 T cells that had entered the FRT 90–120 days after immunization, thus providing an opportunity to test this model. We found that $\sim 30\%$ of CD69[−] migrating memory P14 CD8 T cells in naive parabionts were CD62L⁺, indicating that much of the NLT recirculating population would conventionally be defined as T_{CM} (Figure 6D).

We next tested whether T_{EM} are in fact specialized to migrate to NLT sites of inflammation compared to T_{CM} . CD62L⁺ (T_{CM}) (5×10^5) or CD62L[−] (T_{EM}) (5×10^5) memory OT-I CD8 T cells (see Experimental Procedures) were transferred into P14 immune chimeras. The next day, mice were challenged trans-cervically with gp33 peptide to reactivate P14 T_{RM} in the FRT and precipitate an inflammatory response that recruits circulating memory T cells (Schenkel et al., 2013). As shown in Figure 6E, T_{CM} and T_{EM} migrated to NLT inflammation equivalently, revising the current model of how each subset participates in host immunity.

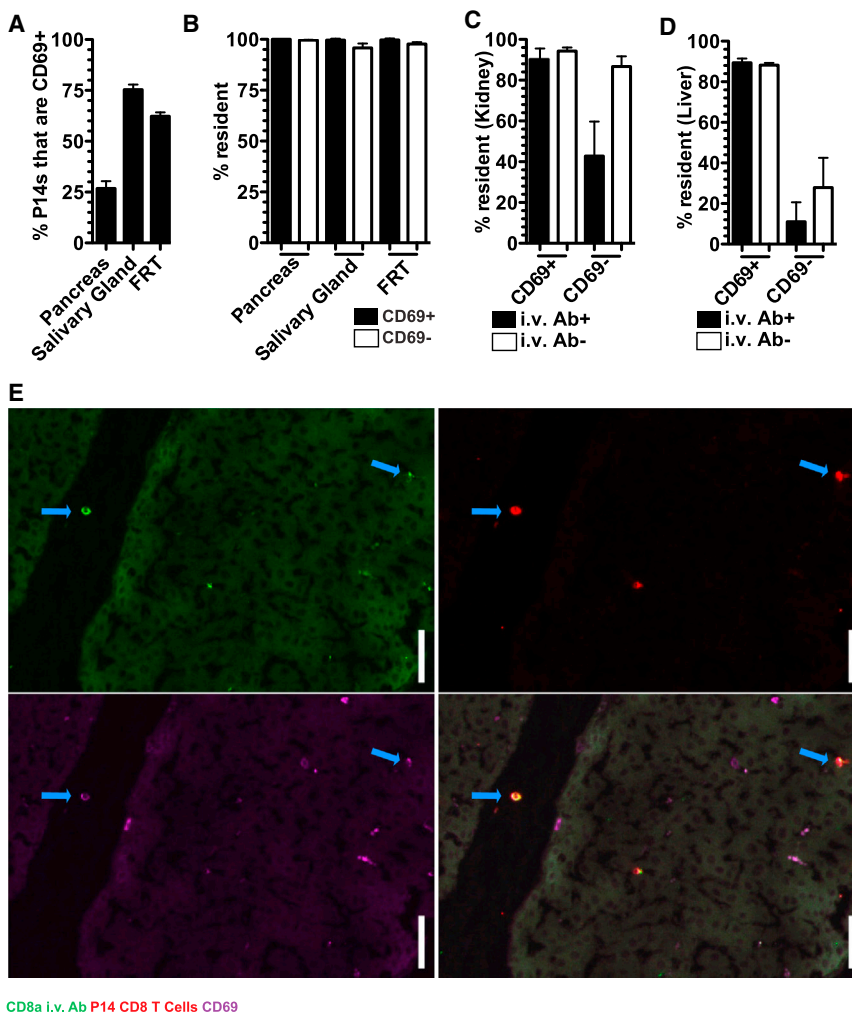


Figure 5. CD69 Is an Imperfect Marker of Tissue Residence

(A) P14 CD8 T cells from immune parabionts were analyzed for the expression of CD69 in the pancreas, salivary gland, and FRT by QIM.

(B) The fraction of CD69⁺ and CD69⁻ P14 CD8 T cells that were resident.

(C and D) The percent of P14 CD8 T cells that were resident among i.v. Ab^{+/+} and CD69^{+/+} in the kidney (C) and (D) liver.

(E) Representative image of a CD69⁺ i.v. Ab⁺ P14 CD8 T cell in a large vessel in the liver. α -CD8 α i.v. Ab (green), P14 CD8 T cells (red), and CD69 (purple). Blue arrows indicate α -CD8 α i.v. Ab⁺ CD69⁺ P14 CD8 T cells. Scale bar, 20 μ m. n = 3, representative of nine mouse pairs from three independent experiments. Graphs show mean and SEM.

See also Figure S2.

et al., 2012; data not shown). In particular, skin has been shown to harbor abundant memory T cells in humans, where extraction efficiency is also an important challenge (Clark et al., 2006). This study further highlights the abundance of T_{RM} as well as their broad anatomic distribution, which includes the BMP. Moreover, based on cell isolation and flow cytometry enumerations, cells in mucosal tissues were 50- to 400-fold more rare than in SLOs. However, QIM revealed that the ratios of memory CD8 T cells relative to potential targets (i.e., host cells) were fairly comparable between SLO and NLT. These observations revise perceptions of immunosurveillance and may help

DISCUSSION

This study provides a rigorous and comprehensive analysis of the anatomic distribution of a single memory CD8 T cell population. Preparation of single cell suspensions from tissues recovered as few as 2% of memory CD8 T cells from NLT and inaccurately represented memory T cell subsets, phenotype, and tissue distribution. Similar results were observed in human tissue, suggesting fundamental errors with standard techniques that we rely upon for our basic characterization of the peripheral immune system. These issues may extend to other hematopoietic lineages, evaluation of vaccine responses in tissues, and other clinical investigations.

When the NLT population was summated with the unexpected abundance of memory CD8 T cells observed in BMP, SLO (WP of spleen and the 37 macroscopic LNs in mice) did not contain the majority of memory CD8 T cells (Van den Broeck et al., 2006). Our study likely underestimates NLT memory CD8 T cells because not every tissue was analyzed, including many other locations (heart, bladder, gall bladder, esophagus, trachea, skeletal muscle, etc.) that contain memory CD8 T cells (Casey

et al., 2012; data not shown). In particular, skin has been shown to harbor abundant memory T cells in humans, where extraction efficiency is also an important challenge (Clark et al., 2006). This study further highlights the abundance of T_{RM} as well as their broad anatomic distribution, which includes the BMP. Moreover, based on cell isolation and flow cytometry enumerations, cells in mucosal tissues were 50- to 400-fold more rare than in SLOs. However, QIM revealed that the ratios of memory CD8 T cells relative to potential targets (i.e., host cells) were fairly comparable between SLO and NLT. These observations revise perceptions of immunosurveillance and may help

explain why frontline memory CD8 T cell populations can rapidly detect infections in barrier tissues (Gebhardt et al., 2009; Jiang et al., 2012; Shin and Iwasaki, 2012; Teijaro et al., 2011; Wu et al., 2014). We focused most analyses on memory resulting from a single infection in order to achieve the depth of characterization described here. However, evidence supports that fundamental observations regarding the abundance of resident memory extend well beyond the context of LCMV. Many infections, whether systemic or local, result in CD8 T cell populations that express peripheral homing molecules and then become broadly distributed throughout multiple nonlymphoid tissues (Masopust et al., 2004, 2010; Liu et al., 2006; Kaufman et al., 2008). In fact, even lymphopenia-induced proliferation is sufficient to induce widespread CD8 T cell dissemination and acquisition of markers associated with T_{RM} (Casey et al., 2012). These data indicate that T_{RM} development may occur irrespective of local antigen or inflammation. T_{RM} are likely not only widely distributed in a variety of contexts, but also underestimated. Indeed, recent evidence suggests that most CD8 T cells that express markers of antigen-experience also express CD69 when isolated from

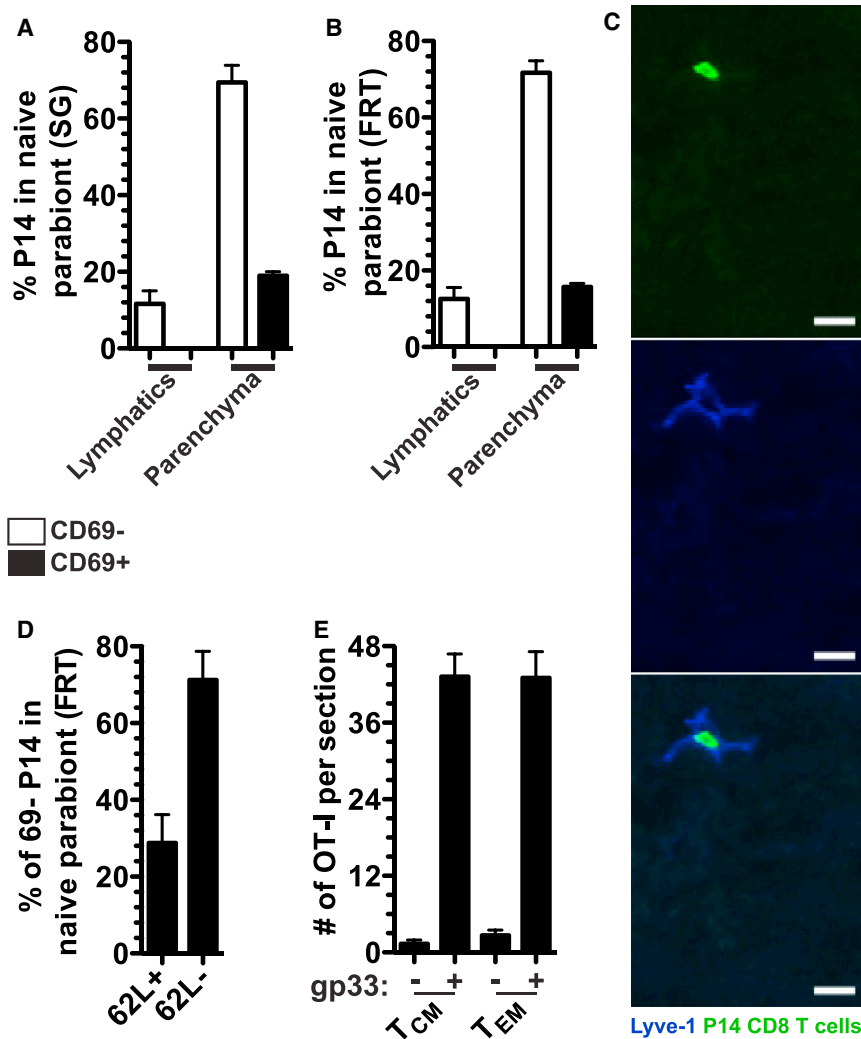


Figure 6. Migration of Memory CD8 T Cell Subsets

(A and B) P14 CD8 T cells analyzed by QIM from naive parabionts were quantified based on their localization within the parenchyma or afferent lymphatic Lyve-1⁺ vessels and for the expression of CD69 in the (A) salivary gland and (B) female reproductive tract.

(C) Representative image of a P14 CD8 T cell in the FRT afferent lymphatics of a naive parabiont. Lyve-1 (blue) and P14 CD8 T cells (green). Scale bar, 10 μ m.

(D) Fraction of CD69⁻ P14 CD8 T cells in the FRT of the naive parabiont that were CD62L⁺ or CD62L⁻. n = 3, representative of nine mice from three independent experiments.

(E) CD62L⁺ (5×10^5) or CD62L⁻ (5×10^5) memory OT-I CD8 T cells isolated from the spleen of VSV-OVA immune chimeras were transferred into P14 immune chimeras and the next day P14 immune chimeras were challenged transcutaneously with 50 μ g gp33 peptide. Two days later, total numbers of OT-I CD8 T cells were enumerated in the FRT. n = 6, representative of two independent experiments. Graphs show mean and SEM. See also Figure S2.

vasculature is not sufficient to reveal the residence status of a CD8 T lymphocyte. Furthermore, we did detect memory CD8 T cells that had entered certain NLT months after putative clearance of infection. While rare, a substantive proportion of these “latecomers” expressed CD69. It is possible that this represents a one-way trip and that T_{RM} are maintained by a slow matriculation of circulating memory CD8 T cells that convert to T_{RM}, upregulating CD69 post-migration.

human tissues, which suggests that most are resident (Thome et al., 2014). We demonstrated that the isolation of CD8 T cells from nonlymphoid tissues was inefficient in both mice and humans, suggesting that memory T cells outside of secondary lymphoid organs are misrepresented regardless of species or pathogen specificity.

This study also raises important caveats with how we define resident and recirculating memory CD8 T cell subsets. CD69 is considered the lineage-defining marker for T_{RM}. It has been shown that CD69 is important for establishing T_{RM} populations in epidermis after HSV-1 infection in mice (Mackay et al., 2013). In accordance with these data, we found that many T_{RM} were CD69⁺. However, we found that many were not. Moreover, expression of another marker often used to identify T_{RM}, CD103, was compartment-specific and most T_{RM} lacked CD103. These data define additional complexity among T_{RM} and suggest that there is more than one subset. Maintenance of CD69⁻ T_{RM} could be mediated by alternative means such as downregulation of KLF2-dependent S1P receptors (Skon et al., 2013). Our data also reveal that anatomic localization outside (or inside)

To what degree do memory CD8 T cells undergo bona fide recirculation through NLT? Leukocytes exit tissues via the afferent lymphatics. Because we identified latecomer memory CD8 T cells in the lymphatics of the FRT and SG, these cells are likely a bona fide NLT recirculating subset in the steady state. In support of this conclusion, this population did not express CD69.

Given the abundance of memory CD8 T cells in the BMP and NLT and the relative paucity of recirculation through NLT, our data raise questions as to whether most T_{EM} truly survey NLT. Perhaps a more likely scenario is that NLT are surveyed by only a fraction of specialized T_{EM}, and other T_{EM} serve functions that remain to be fully elucidated. Our data indicate that T_{CM} also contribute to the rare population of NLT recirculating memory CD8 T cells in the steady state, which may also occur in human skin (Clark et al., 2006). Moreover, in the context of inflammation, T_{CM} migrated just as robustly as T_{EM} to the FRT. In contrast to the original and elegant T_{CM}/T_{EM} model, this may ensure that there is a long-lived pool capable of being recruited because T_{CM} may be maintained longer than CD62L⁻ BMP (Wherry et al., 2003; Marzo et al., 2005).

Figure S2 summarizes and contextualizes these observations. Most host cells, which require contact by CD8 T cells for immunosurveillance, are positioned outside of secondary lymphoid organs. These include solid organs and body surfaces such as the gastrointestinal, respiratory, and genitourinary mucosae and skin that represent common primary sites of pathogen exposure. The majority of memory CD8 T cells that patrol these frontlines are segregated populations that confine their surveillance locally and do not migrate between other NLT, SLOs, or blood. Therefore, this major fraction of the memory CD8 T cell pool cannot be captured by sampling blood or SLOs. Indeed, the recirculating populations, which included both CD62L⁺ T_{EM} and CD62L⁺ T_{CM}, actually comprised a small minority of those cells patrolling NLT. The blood and marginated pool (BMP) (that includes peripheral blood, the red pulp of the spleen, and vascular compartments within organs such as liver and kidney) also contains a substantial fraction of the overall memory CD8 T cell population. When NLT re-infections are not rapidly eliminated, inflammation recruits both T_{EM} and T_{CM} from the BMP, presumably to contribute to local immunosurveillance and pathogen control. The vascular compartments of certain tissues, including liver and kidney, are also populated by T_{RM}, which may facilitate direct immunosurveillance of the organ via the endothelium, for instance of hepatocytes through sinusoidal fenestrae, or may prevent hematogenous spread of target cells. When infections are not contained within NLTs, pathogens and associated foreign antigens reach the SLOs. Here, T_{CM} (that recirculate between blood and SLOs) can be reactivated to proliferate and provide additional reinforcements that migrate to NLTs.

This revised model highlights the provincial nature of memory CD8 T cell-mediated immunosurveillance. Different populations of memory CD8 T cells patrol distinct anatomic niches that form an integrated immunological network to protect the host in the event of reinfection. However, the majority of the host is patrolled by abundant yet discrete regionalized memory CD8 T cell populations that do not recirculate and instead remain confined within single anatomic compartments.

EXPERIMENTAL PROCEDURES

See also the [Extended Experimental Procedures](#).

Mice, Adoptive Transfers, Surgeries, and Infections

All mice were used in accordance with the Institutional Animal Care and Use Committee at the University of Minnesota. C57BL/6J mice were purchased from The Jackson Laboratory. P14 and OT-I CD8 T cell transgenic mice were maintained in house. P14 immune chimeras were generated by transferring 5×10^4 P14 CD8 T cells into naive C57BL/6J mice. The following day, these mice were infected with 2×10^5 plaque-forming units (PFU) LCMV Armstrong via intraperitoneal (i.p.) injection. For endogenous studies, naive C57BL/6J mice were infected with 2×10^5 PFU LCMV Armstrong i.p. OT-I immune chimeras were generated by transferring 5×10^4 naive OT-I CD8 T cells into C57BL/6 mice. The next day, mice were infected with 2×10^6 PFU Vaccinia Virus expressing chicken ovalbumin. Sixty days after infection, CD62L⁺ and CD62L⁺ memory OT-I splenocytes were purified using α -CD62L PE and α -PE magnetic beads according to the manufacturer's instructions (Miltenyi). CD62L⁺ (5×10^5) or CD62L⁺ (5×10^5) OT-I cells were transferred into P14 immune chimeras that 60 days previously had been infected with LCMV. The following day animals were transcutaneously (t.c.) challenged with

50 μ g gp-33 peptide as previously described (Collins et al., 2009; Schenkel et al., 2013). Parabiosis surgeries were performed as previously described (Schenkel et al., 2013).

Intravascular Antibody

To label all CD8 T cells in compartments contiguous with vasculature, animals were injected i.v. with 3 μ g α -CD8 α biotinylated antibody (53-6.7, eBioscience) that was allowed to circulate for three minutes prior to sacrifice. For detection of i.v.-injected α -CD8 α antibody, fluorochrome-conjugated streptavidin (eBioscience) was used for flow cytometry and donkey anti-rat antibodies (Jackson Laboratory) were used for immunofluorescence.

Isolations and Flow Cytometry

Three minutes after in vivo intravascular antibody injection (Anderson et al., 2014), mice were sacrificed and organs of interest were excised. For isolation of SI IELs, the small intestine was removed, Peyer's patches were excised, and the intestine was cut longitudinally and then laterally into 0.5–1 cm² pieces. Large intestines and stomachs were cut similarly. To remove IELs, small intestine, large intestine, and stomach pieces were incubated with 0.154 mg/ml dithioerythritol (DTE) in 10% HBSS/HEPES bicarbonate for 30 min at 37°C, stirring at 450 rpm. Following IEL isolation, small intestine, large intestine, and stomach pieces were further processed to remove lamina propria lymphocytes (LPL), by treatment with 100 U/ml type I collagenase (Worthington) in RPMI 1640, 5% FBS, 2 mM MgCl₂, 2 mM CaCl₂ for 45 min at 37°C, stirring at 450 rpm. The following tissues were cut into pieces and enzymatically digested with 100 U/ml type I collagenase (Worthington) in RPMI 1640, 5% FBS, 2 mM MgCl₂, 2 mM CaCl₂ at 37°C, stirring at 450 rpm; salivary gland (SG) (mucous portion removed, treated for 45 min), kidney (treated for 45 min), pancreas (treated for 20 min), and lung (treated for 1 hr). For isolation of the female reproductive tract, the uterine horns, cervix, and vaginal tissue were resected and cut into small pieces prior to treatment with 0.5 mg/ml type IV collagenase (Sigma) RPMI 1640, 5% FBS, 2 mM MgCl₂, 2 mM CaCl₂ (treated for 1 hr) at 37°C, stirring at 450 rpm. After enzymatic treatment, the remaining tissue pieces of the stomach LPL, FRT, SG, pancreas, lung, and kidney were further mechanically disrupted by a gentleMACS Dissociator (setting m_Spleen_01.01). The liver was mechanically dissociated using the back of a syringe over a 70- μ m nylon cell strainer (Falcon). From single cell suspensions, lymphocytes were separated using a 44/67% Percoll density gradient. Spleen, lymph nodes, and thymus were mechanically dissociated using the back of a syringe against a polystyrene Petri dish that had previously been scored in four directions with an 18.5 gauge needle. Peripheral blood was treated with ACK lysis buffer. The resulting single cell suspension was stained for acquisition on an LSR II flow cytometer (BD Biosciences).

The following antibodies were used for flow cytometry of mouse cells: α -CD103 (M290) from BD Biosciences; α -CD8 α (53-6.7), α -Thy1.1 (HIS51), α -CD44 (IM7), Streptavidin APC, and α -CD45.1 (A20) from eBioscience; and α -Thy1.1 (OX-7) and α -CD8 β (YTS156.7.7) from Biolegend.

Quantitative Immunofluorescence Microscopy

To determine volumes of individual organs, mice age-matched to those analyzed for enumeration were sacrificed, and organs were removed and cleared of all fat, connective tissue, and fecal matter. Each organ was submerged in PBS, the displaced volume was measured, and this was repeated for each organ four times. This displacement procedure was conducted on six mice age-matched to those used in experiments. For organs too small for accurate volume displacement, including the mandibular lymph nodes, organs were pooled from multiple animals before measuring displacement and dividing the displaced volume by the number of pooled organs. For QIM enumeration, 3 min after in vivo intravascular antibody injection, mice were sacrificed and organs of interest were excised, positioned in plastic cryomolds and snap frozen in optimum cutting temperature (O.C.T.) freezing medium. From these frozen tissue blocks, slides of 7- μ m sections were prepared. Slides were stained for acquisition on a Leica DM5500B 4 color fluorescent system with motorized z focus stage for fully automated image stitching. Enumeration of P14 cells as well as CD103, CD69, and CD62L expression was done manually in Adobe Photoshop. ImageJ64 software was used to enumerate

nuclei in each image (as stained by DAPI) as previously described (Schenkel et al., 2013), all counts were manually validated, and these, like the manual enumerations, were extrapolated to whole organs. Area measurements of images were made either in LAS (Leica Acquisition Software) or Adobe Photoshop. Area measurements were multiplied by tissue section thickness (7- μ m) to determine the volume of enumerated images. Manual and ImageJ64 counts were extrapolated up to whole organ enumerations. We multiplied all enumerations by 11/19 to correct for all cells that would be counted twice because they straddle two adjacent sections. This correction factor is derived because sections are 7 μ m thick, the diameter of a memory CD8 T lymphocyte is \sim 7 μ m, and any cell traversing a section by >1 μ m would be enumerated (Decoursey et al., 1987). Sections through whole organs or large (\sim 5 mm²) tiled images were counted, no fewer than 100 and up to 3,000 P14 were counted per organ per animal, representative tissue sections were sampled that included diverse regions of each organ and non-serial sections (35–70 μ m apart) to ensure P14 counts were representative of the entire organ. For example, whole sections of the stomach were counted to ensure anatomical representation of the fundus, body, and antrum regions. The following antibodies were used for immunofluorescence microscopy: α -CD103 (2E7) and α -Thy1.1 (OX-7) from Biolegend; α -CD62L (MEL-14), α -CD8 α (53-6.7), α -CD8 β (YTS1 56.7.7), α -E-cadherin (DECMA-1), α -CD45.1 (A20) from eBioscience; α -CD69 (polyclonal goat), α -Lyve-1(223322) from R&D; α -Cytokeratin 8 (rabbit polyclonal), α -Cytokeratin 18 (rabbit polyclonal), α -PE (rabbit polyclonal) from Novus Biologicals; α -Collagen IV (goat polyclonal) from Millipore; and α -Cytokeratin5 (PRB-160P) from Covance. DAPI and prolong gold were from Invitrogen. The following secondary antibodies were from Jackson ImmunoResearch: donkey α -rabbit (polyclonal), bovine α -goat (polyclonal), and donkey α -rat (polyclonal).

DNA Extraction

To validate QIM extrapolation, DNA content of whole organs was determined. First organs were dissected, cut into 1-mm pieces, and digested in tissue digestion buffer (10 mM TRIS, 10 mM EDTA, 10% SDS, sodium acetate and proteinase K) shaking overnight at 56°C. Phenol-chloroform-isoamyl alcohol DNA extraction was then performed on each digested organ. Each DNA sample was resuspended in TE buffer and nucleic acid concentration was determined by a nanodrop spectrophotometer. Each sample was measured 4 times; an average of the four was taken to determine the most accurate nucleic acid content of each sample. The total nucleic acid content of each organ was divided by an assumed 6 pg of DNA per cell to determine total cell number for the organ based on DNA content (dos Anjos Pires et al., 2001).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures and two figures and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.03.031>.

AUTHOR CONTRIBUTIONS

E.M.S., J.M.S., L.S.M., and D.M. conceived and designed the experiments. E.M.S., J.M.S., L.K.B., K.A.F., and B.Z.I. performed the experiments. L.S.M. and P.J.S. contributed reagents/materials/analysis tools. E.M.S., J.M.S., and D.M. wrote the manuscript.

ACKNOWLEDGMENTS

This work was supported by grants from the NIH (R01-AI084913 and R01-AI111671 to D.M., T32-AI083196 to E.M.S., F30-DK100159-02 to J.M.S., and F31-CA183226 to L.S.M.).

Received: September 19, 2014

Revised: December 9, 2014

Accepted: February 23, 2015

Published: May 7, 2015

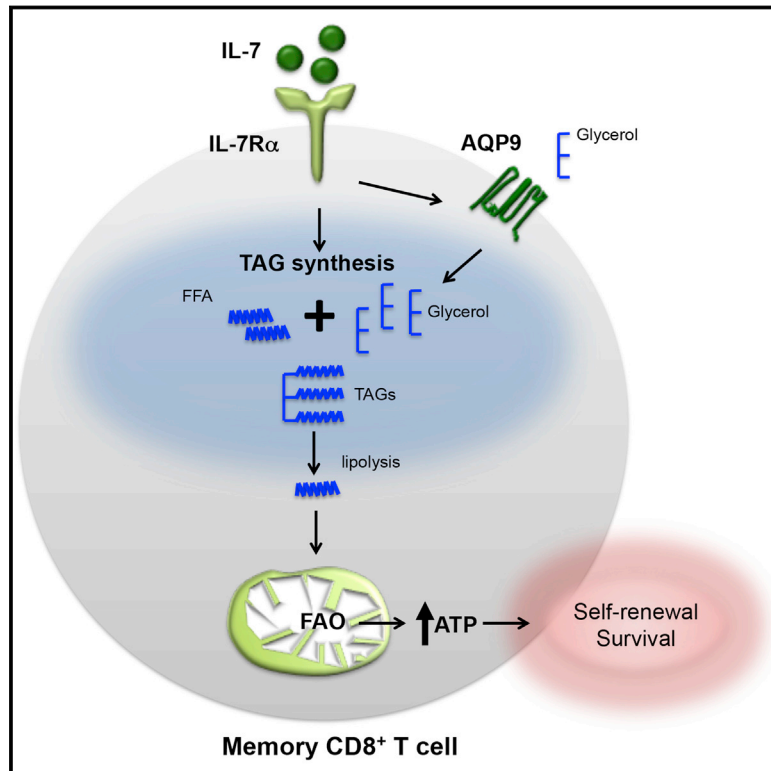
REFERENCES

- Altman, J.D., Moss, P.A., Goulder, P.J., Barouch, D.H., McHeyzer-Williams, M.G., Bell, J.I., McMichael, A.J., and Davis, M.M. (1996). Phenotypic analysis of antigen-specific T lymphocytes. *Science* 274, 94–96.
- Anderson, K.G., Mayer-Barber, K., Sung, H., Beura, L., James, B.R., Taylor, J.J., Qunaj, L., Griffith, T.S., Vezys, V., Barber, D.L., and Masopust, D. (2014). Intravascular staining for discrimination of vascular and tissue leukocytes. *Nat. Protoc.* 9, 209–222.
- Casey, K.A., Fraser, K.A., Schenkel, J.M., Moran, A., Abt, M.C., Beura, L.K., Lucas, P.J., Artis, D., Wherry, E.J., Hogquist, K., et al. (2012). Antigen-independent differentiation and maintenance of effector-like resident memory T cells in tissues. *J. Immunol.* 188, 4866–4875.
- Clark, R.A., Chong, B., Mirchandani, N., Brinster, N.K., Yamanaka, K., Dorgiet, R.K., and Kupper, T.S. (2006). The vast majority of CLA+ T cells are resident in normal skin. *J. Immunol.* 176, 4431–4439.
- Collins, M.K., Tay, C.-S., and Erlebacher, A. (2009). Dendritic cell entrapment within the pregnant uterus inhibits immune surveillance of the maternal/fetal interface in mice. *J. Clin. Invest.* 119, 2062–2073.
- Decoursey, T.E., Chandy, K.G., Gupta, S., and Cahalan, M.D. (1987). Mitogen induction of ion channels in murine T lymphocytes. *J. Gen. Physiol.* 89, 405–420.
- Doctor, R.B., Serkova, N.J., Hasebroock, K.M., Zafar, I., and Edelstein, C.L. (2010). Distinct patterns of kidney and liver cyst growth in pkd2(WS25/-) mice. *Nephrol. Dial. Transplant.* 25, 3496–3504.
- dos Anjos Pires, M., Palmeira, C., Rodrigues, P., Lopes, C., and Oliveira-Torres, F. (2001). Establishment of a diploid reference value for DNA ploidy analysis by image cytometry in mouse cells. *Anal. Quant. Cytol. Histol.* 23, 427–432.
- Farber, D.L., Yudanin, N.A., and Restifo, N.P. (2014). Human memory T cells: generation, compartmentalization and homeostasis. *Nat. Rev. Immunol.* 14, 24–35.
- Galkina, E., Thattai, J., Dabak, V., Williams, M.B., Ley, K., and Braciale, T.J. (2005). Preferential migration of effector CD8+ T cells into the interstitium of the normal lung. *J. Clin. Invest.* 115, 3473–3483.
- Ganusev, V.V., and De Boer, R.J. (2007). Do most lymphocytes in humans really reside in the gut? *Trends Immunol.* 28, 514–518.
- Gebhardt, T., Wakim, L.M., Eidsmo, L., Reading, P.C., Heath, W.R., and Carbone, F.R. (2009). Memory T cells in nonlymphoid tissue that provide enhanced local immunity during infection with herpes simplex virus. *Nat. Immunol.* 10, 524–530.
- Horbul, J.E., Schmechel, S.C., Miller, B.R., Rice, S.A., and Southern, P.J. (2011). Herpes simplex virus-induced epithelial damage and susceptibility to human immunodeficiency virus type 1 infection in human cervical organ culture. *PLoS ONE* 6, e22638. <http://dx.doi.org/10.1371/journal.pone.0022638>.
- Jiang, X., Clark, R.A., Liu, L., Wagers, A.J., Fuhlbrigge, R.C., and Kupper, T.S. (2012). Skin infection generates non-migratory memory CD8+ T(RM) cells providing global skin immunity. *Nature* 483, 227–231.
- Kaufman, D.R., Liu, J., Carville, A., Mansfield, K.G., Havenga, M.J.E., Goudsmit, J., and Barouch, D.H. (2008). Trafficking of antigen-specific CD8+ T lymphocytes to mucosal surfaces following intramuscular vaccination. *J. Immunol.* 181, 4188–4198.
- Khanna, K.M., McNamara, J.T., and Lefrançois, L. (2007). In situ imaging of the endogenous CD8 T cell response to infection. *Science* 318, 116–120.
- Klonowski, K.D., Williams, K.J., Marzo, A.L., Blair, D.A., Lingenheld, E.G., and Lefrançois, L. (2004). Dynamics of blood-borne CD8 memory T cell migration in vivo. *Immunity* 20, 551–562.
- Lee, Y.T., Suarez-Ramirez, J.E., Wu, T., Redman, J.M., Bouchard, K., Hadley, G.A., and Cauley, L.S. (2011). Environmental and antigen receptor-derived signals support sustained surveillance of the lungs by pathogen-specific cytotoxic T lymphocytes. *J. Virol.* 85, 4085–4094.

- Liu, L., Fuhlbrigge, R.C., Karibian, K., Tian, T., and Kupper, T.S. (2006). Dynamic programming of CD8⁺ T cell trafficking after live viral immunization. *Immunity* 25, 511–520.
- Mackay, L.K., Rahimpour, A., Ma, J.Z., Collins, N., Stock, A.T., Hafon, M.-L., Vega-Ramos, J., Lauzurica, P., Mueller, S.N., Stefanovic, T., et al. (2013). The developmental pathway for CD103⁺CD8⁺ tissue-resident memory T cells of skin. *Nat. Immunol.* 14, 1294–1301.
- Marzo, A.L., Klonowski, K.D., Le Bon, A., Borrow, P., Tough, D.F., and Lefrançois, L. (2005). Initial T cell frequency dictates memory CD8⁺ T cell lineage commitment. *Nat. Immunol.* 6, 793–799.
- Masopust, D., and Schenkel, J.M. (2013). The integration of T cell migration, differentiation and function. *Nat. Rev. Immunol.* 13, 309–320.
- Masopust, D., Vezys, V., Marzo, A.L., and Lefrançois, L. (2001). Preferential localization of effector memory cells in nonlymphoid tissue. *Science* 291, 2413–2417.
- Masopust, D., Vezys, V., Usherwood, E.J., Cauley, L.S., Olson, S., Marzo, A.L., Ward, R.L., Woodland, D.L., and Lefrançois, L. (2004). Activated primary and memory CD8 T cells migrate to nonlymphoid tissues regardless of site of activation or tissue of origin. *J. Immunol.* 172, 4875–4882.
- Masopust, D., Vezys, V., Wherry, E.J., Barber, D.L., and Ahmed, R. (2006). Cutting edge: gut microenvironment promotes differentiation of a unique memory CD8 T cell population. *J. Immunol.* 176, 2079–2083.
- Masopust, D., Choo, D., Vezys, V., Wherry, E.J., Duraiswamy, J., Akondy, R., Wang, J., Casey, K.A., Barber, D.L., Kawamura, K.S., et al. (2010). Dynamic T cell migration program provides resident memory within intestinal epithelium. *J. Exp. Med.* 207, 553–564.
- Mueller, S.N., Gebhardt, T., Carbone, F.R., and Heath, W.R. (2013). Memory T cell subsets, migration patterns, and tissue residence. *Annu. Rev. Immunol.* 31, 137–161.
- Murali-Krishna, K., Altman, J.D., Suresh, M., Sourdiv, D.J., Zajac, A.J., Miller, J.D., Slansky, J., and Ahmed, R. (1998). Counting antigen-specific CD8 T cells: a reevaluation of bystander activation during viral infection. *Immunity* 8, 177–187.
- Nakanishi, Y., Lu, B., Gerard, C., and Iwasaki, A. (2009). CD8⁺ T lymphocyte mobilization to virus-infected tissue requires CD4⁺ T-cell help. *Nature* 462, 510–513.
- Nutter, R.L., Gridley, D.S., Slater, J.M., and McMillan, P.J. (1980). Responses of mouse spleen morphology to the growth of subcutaneously injected virally transformed cells. *Anat. Rec.* 197, 363–368.
- Peaudecerf, L., and Rocha, B. (2011). Role of the gut as a primary lymphoid organ. *Immunol. Lett.* 140, 1–6.
- Reinhardt, R.L., Khoruts, A., Merica, R., Zell, T., and Jenkins, M.K. (2001). Visualizing the generation of memory CD4 T cells in the whole body. *Nature* 410, 101–105.
- Rocha, B., Vassalli, P., and Guy-Grand, D. (1991). The V beta repertoire of mouse gut homodimeric alpha CD8⁺ intraepithelial T cell receptor alpha/beta + lymphocytes reveals a major extrathymic pathway of T cell differentiation. *J. Exp. Med.* 173, 483–486.
- Sallusto, F., Lenig, D., Förster, R., Lipp, M., and Lanzavecchia, A. (1999). Two subsets of memory T lymphocytes with distinct homing potentials and effector functions. *Nature* 401, 708–712.
- Scheller, L.F., Wirtz, R.A., and Azad, A.F. (1994). Susceptibility of different strains of mice to hepatic infection with *Plasmodium berghei*. *Infect. Immun.* 62, 4844–4847.
- Schenkel, J.M., Fraser, K.A., Vezys, V., and Masopust, D. (2013). Sensing and alarm function of resident memory CD8⁺ T cells. *Nat. Immunol.* 14, 509–513.
- Schenkel, J.M., Fraser, K.A., and Masopust, D. (2014). Cutting edge: resident memory CD8 T cells occupy frontline niches in secondary lymphoid organs. *J. Immunol.* 192, 2961–2964.
- Selby, W.S., Janossy, G., Bofill, M., and Jewell, D.P. (1984). Intestinal lymphocyte subpopulations in inflammatory bowel disease: an analysis by immunohistological and cell isolation techniques. *Gut* 25, 32–40.
- Shin, H., and Iwasaki, A. (2012). A vaccine strategy that protects against genital herpes by establishing local memory T cells. *Nature* 491, 463–467.
- Skon, C.N., Lee, J.Y., Anderson, K.G., Masopust, D., Hogquist, K.A., and Jameson, S.C. (2013). Transcriptional downregulation of S1pr1 is required for the establishment of resident memory CD8⁺ T cells. *Nat. Immunol.* 14, 1285–1293.
- Suvas, P.K., Dech, H.M., Sambira, F., Zeng, J., and Onami, T.M. (2007). Systemic and mucosal infection program protective memory CD8 T cells in the vaginal mucosa. *J. Immunol.* 179, 8122–8127.
- Teijaro, J.R., Turner, D., Pham, Q., Wherry, E.J., Lefrançois, L., and Farber, D.L. (2011). Cutting edge: Tissue-retentive lung memory CD4 T cells mediate optimal protection to respiratory virus infection. *J. Immunol.* 187, 5510–5514.
- Thome, J.J.C., Yudanin, N., Ohmura, Y., Kubota, M., Grinshpun, B., Sathaliyawa, T., Kato, T., Lerner, H., Shen, Y., and Farber, D.L. (2014). Spatial map of human T cell compartmentalization and maintenance over decades of life. *Cell* 159, 814–828.
- Van den Broeck, W., Derore, A., and Simoons, P. (2006). Anatomy and nomenclature of murine lymph nodes: Descriptive study and nomenclatory standardization in BALB/cAnNCrI mice. *J. Immunol. Methods* 312, 12–19.
- Wherry, E.J., Teichgräber, V., Becker, T.C., Masopust, D., Kaech, S.M., Antia, R., von Andrian, U.H., and Ahmed, R. (2003). Lineage relationship and protective immunity of memory CD8 T cell subsets. *Nat. Immunol.* 4, 225–234.
- Wu, T., Hu, Y., Lee, Y.-T., Bouchard, K.R., Benechet, A., Khanna, K., and Cauley, L.S. (2014). Lung-resident memory CD8 T cells (T_{RM}) are indispensable for optimal cross-protection against pulmonary virus infection. *J. Leukoc. Biol.* 95, 215–224.

IL-7-Induced Glycerol Transport and TAG Synthesis Promotes Memory CD8⁺ T Cell Longevity

Graphical Abstract



Authors

Guoliang Cui, Matthew M. Staron, ..., Jingxia Wu, Susan M. Kaech

Correspondence

susan.kaech@yale.edu

In Brief

Interleukin-7 induces expression of the glycerol channel aquaporin 9, allowing memory CD8⁺ T cells to import glycerol, use it for triglyceride synthesis and storage, and sustain ATP levels required for long-term metabolic fitness and fast responses to reinfection.

Highlights

- IL-7 induces glycerol channel AQP9 expression in CD8⁺ T cells
- AQP9 is required for memory CD8⁺ T cell survival and self-renewal
- AQP9 imports glycerol, promotes TAG synthesis, and sustains ATP levels in T cells
- IL-7 enhances TAG synthesis to promote memory CD8⁺ T cell survival



IL-7-Induced Glycerol Transport and TAG Synthesis Promotes Memory CD8⁺ T Cell Longevity

Guoliang Cui,¹ Matthew M. Staron,¹ Simon M. Gray,¹ Ping-Chih Ho,¹ Robert A. Amezcua,^{1,2} Jingxia Wu,³ and Susan M. Kaech^{1,2,*}

¹Department of Immunobiology, Yale University School of Medicine, New Haven, CT 06520, USA

²Howard Hughes Medical Institute, 4000 Jones Bridge Road, Chevy Chase, MD 20815-6789, USA

³Department of Internal Medicine, Yale University School of Medicine, New Haven, CT 06520, USA

*Correspondence: susan.kaech@yale.edu

<http://dx.doi.org/10.1016/j.cell.2015.03.021>

SUMMARY

Memory T cells are critical for long-term immunity against reinfection and require interleukin-7 (IL-7), but the mechanisms by which IL-7 controls memory T cell survival, particularly metabolic fitness, remain elusive. We discover that IL-7 induces expression of the glycerol channel aquaporin 9 (AQP9) in virus-specific memory CD8⁺ T cells, but not naive cells, and that AQP9 is vitally required for their long-term survival. AQP9 deficiency impairs glycerol import into memory CD8⁺ T cells for fatty acid esterification and triglyceride (TAG) synthesis and storage. These defects can be rescued by ectopic expression of TAG synthases, which restores lipid stores and memory T cell survival. Finally, we find that TAG synthesis is a central component of IL-7-mediated survival of human and mouse memory CD8⁺ T cells. This study uncovers the metabolic mechanisms by which IL-7 tailors the metabolism of memory T cells to promote their longevity and fast response to rechallenge.

INTRODUCTION

Immunological memory is the foundation of protective vaccines, and therefore, understanding how memory lymphocytes form and persist after vaccination or infection is of great clinical importance. During acute viral infections, antigen-specific CD8⁺ T cells undergo clonal expansion and differentiate into effector T cells that help fight off invading pathogens. After pathogen clearance, the majority of effector cells die and a small population survives as memory T cells, which can be further categorized into central memory T cells (T_{CM}), effector memory T cells (T_{EM}), and tissue resident memory T cells (T_{RM}) based on different migratory and functional properties (Beura and Masopust, 2014). Memory T cells can persist for decades and their longevity in many tissues is dependent on the cytokines IL-7 and IL-15, which promote cell survival and self-renewal (Becker et al., 2002; Kaech et al., 2003; Kennedy et al., 2000; Kieper et al., 2002; Kondrack et al., 2003; Lenz et al., 2004; Schluns et al., 2000). Voluminous

evidence indicates that IL-7 plays an essential role in lymphopoiesis and peripheral T cell survival (Peschon et al., 1994; von Freuden-Jeffry et al., 1995), and our current understanding is that IL-7 promotes survival of naive and memory T cells as well as thymocytes through sustained expression of the anti-apoptotic factors Bcl-2 and Mcl1 (Opferman et al., 2003; Rathmell et al., 2001). However, other IL-7-dependent cellular processes are involved because Bcl-2 overexpression or deletion of Bim or Bax is insufficient to fully rescue T cell development in IL-7 receptor alpha (IL-7R α)-deficient mice (Akashi et al., 1997; Khaled et al., 2002; Maraskovsky et al., 1997; Pellegrini et al., 2004). Indeed, IL-7 also controls amino acids uptake and glucose utilization in normal and leukemic T cells via its ability to enhance Glut1 trafficking and glycolysis through signal transducer and activator of transcription 5 (STAT5) and AKT activation (Barata et al., 2004; Pearson et al., 2012; Wofford et al., 2008). However, it is not known if IL-7 controls other processes essential for long-term survival of memory T cells nor how naive and memory T cells, which both rely on IL-7, avoid competition with one other for this limited resource.

Recent studies have suggested that a metabolic switch accompanies the differentiation of memory CD8⁺ T cells from activated effector cells. After viral clearance, effector T cells that were once performing high rates of aerobic glycolysis, glutaminolysis, and anabolic metabolism rest down and become more reliant on fatty acid oxidation (FAO) and mitochondrial oxidative phosphorylation (OXPHOS) to generate energy (Fox et al., 2005; Pearce et al., 2009). In support of this model, knock down of lysosomal acid lipase (LAL), an enzyme that releases FAs from triacylglyceride (TAG)s in the lysosome, or carnitine palmitoyl-transferase 1a (CPT1a), an enzyme required for mitochondrial FA transport, suppresses FAO and memory T cell survival following infection (van der Windt et al., 2012). Interestingly, at steady state, memory CD8⁺ T cells do not display high rates of FA uptake, as opposed to activated T cells (O'Sullivan et al., 2014), and therefore, it is not known how these cells maintain an ample supply of FAs over long periods of time to sustain lipid burning. Most cell types, particularly adipocytes, store FAs in the form of TAGs by esterifying three FA chains to glycerol, which can then be broken down to supply FAs for FAO to meet energy demands (Lass et al., 2011).

To better understand the metabolic control of memory CD8⁺ T cell longevity and homeostasis, we profiled the expression of genes involved in cellular metabolism as CD8⁺ T cells

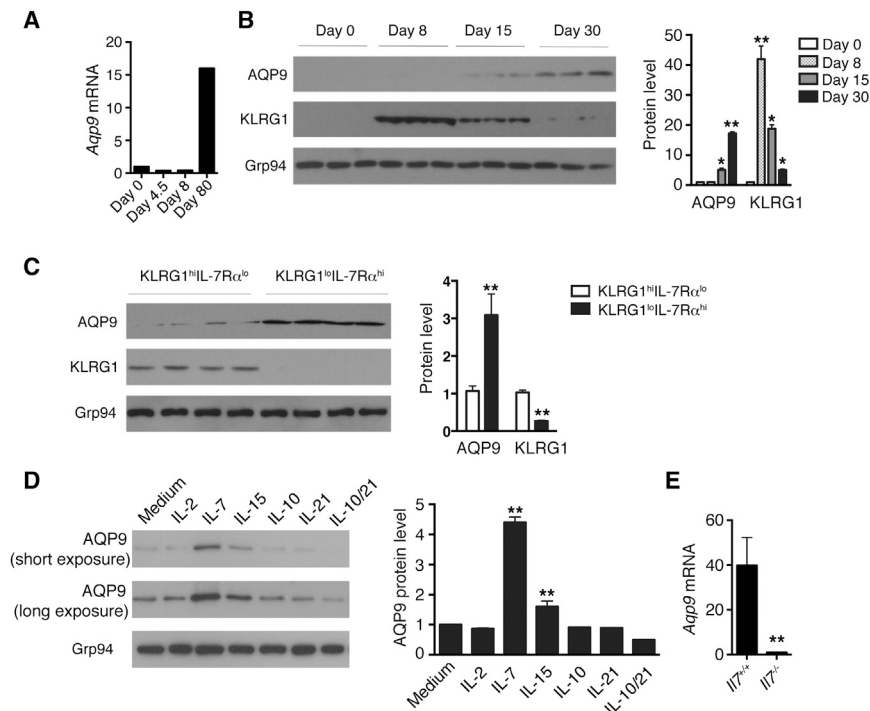


Figure 1. IL-7 Induces AQP9 Expression Selectively in Anti-viral Memory CD8⁺ T Cells and Their Precursors

(A–C) Naive, effector, and memory P14 CD8⁺ T cells were purified on the indicated dpi and the amount of *Aqp9* mRNA was measured using DNA microarrays and analyzed by GeneSpring software (A) or protein using western blotting (B and C).

(A) mRNA is normalized to naive samples.

(B and C) Each lane represents an individual biological sample; Grp94 and KLRG1 were used as loading and internal monitoring controls, respectively.

(C) KLRG1^{hi}IL-7R^{lo} and KLRG1^{lo}IL-7R^{hi} effector CD8⁺ T cell subsets were isolated at 14 dpi. The bar graphs on the right show densitometry quantification of the immunoblot bands.

(D) P14 CD8⁺ T cells were primed with GP_{33–41} peptide for 3 days and then stimulated with various cytokines as indicated for 3 days before western blotting for AQP9. The bar graph on the right shows densitometry quantification of the immunoblot bands.

(E) In vitro primed P14 CD8⁺ T cells were transferred to *I17*^{+/+} or *I17*^{–/–} mice. At 7 days later, the donor cells were purified and *Aqp9* mRNA levels were measured by qRT-PCR.

Data in (B)–(E) are representative of two independent experiments (n = 3–6 mice/group): *p < 0.05 and **p < 0.01 (see also Figure S1).

differentiate from naive → effector → memory stages. This identified that AQP9, a critical glycerol channel in mammals (Carbrey et al., 2003; Rojek et al., 2007), was selectively expressed in CD8⁺ memory T cells compared with naive and effector T cells. Through biochemical and genetic analyses, we found that IL-7 induced AQP9 expression, glycerol importation, and TAG synthesis, which was necessary for memory CD8⁺ T cell survival and homeostasis. Thus, this study reveals a previously unknown metabolic role for IL-7 in directing glycerol uptake and TAG storage to sustain memory CD8⁺ T cells long-term survival, and identifies TAG synthesis as a critical biochemical process for therapeutic modulation of memory T cell survival and self-renewal.

RESULTS

IL-7 Induces AQP9 Expression in Memory CD8⁺ T Cells

Aqp9 has a unique temporal gene expression pattern in virus-specific CD8⁺ T cells, shared with only a handful of other genes, being expressed at very low levels in virus-specific naive and effector T cells and progressively increasing as memory T cells form following viral infection (Best et al., 2013) (Figure 1A). Consistent with the *Aqp9* mRNA expression pattern, AQP9 protein was more abundant in memory CD8⁺ T cells (30 days post infection [dpi]) than in naive or effector CD8⁺ T cells isolated 8 and 15 dpi (Figure 1B). Conversely, the expression of KLRG1, an inhibitory receptor expressed on the most terminally differentiated effector CD8⁺ T cells, declined as these cells waned over time (Joshi et al., 2007; Kaech et al., 2003). Further fractionation of the effector CD8⁺ T cells into KLRG1^{hi} IL-7R^{lo} terminal

effector and KLRG1^{lo} IL-7R^{hi} memory precursor effector cell subsets at 14 dpi revealed that AQP9 was selectively expressed in the IL-7R^{hi} effector cells that preferentially seed the memory T cell pool (Figure 1C) (Kaech et al., 2003; Schluns et al., 2000).

To determine if cytokines that regulate effector and memory CD8⁺ T cell development and homeostasis induce AQP9 in CD8⁺ T cells, we stimulated lymphocytic choriomeningitis virus (LCMV)-specific P14 TCR tg CD8⁺ T cells, which recognize the LCMV epitope GP_{33–41}, in vitro with peptide for three days and then with IL-2, IL-7, IL-15, IL-10, and IL-21 and examined AQP9 expression using western blotting three days later. This showed that IL-7, and to a lesser extent IL-15, induced AQP9 expression in activated CD8⁺ T cells (Figure 1D). To further test the requirement of IL-7 for AQP9 expression in an in vivo setting, we transferred in vitro primed P14 CD8⁺ T cells to *I17*^{+/+} or *I17*^{–/–} mice and analyzed *Aqp9* expression in the cells 7 days later using quantitative (q)RT-PCR. This showed that *Aqp9* mRNA was dramatically decreased in cells isolated from *I17*^{–/–} host mice, indicating that IL-7 signaling was both necessary and sufficient to sustain *Aqp9* expression in antigen-experienced CD8⁺ T cells (Figure 1E).

AQP9 Deficiency Impairs Memory CD8⁺ T Cell Survival following Infection

The observation that AQP9 was selectively expressed in mature memory CD8⁺ T cells following acute viral infection prompted us to examine its functional role in memory T cell generation. To this end, we generated 50:50 mixed *Aqp9*^{+/+} and *Aqp9*^{–/–} bone marrow chimeric mice. At eight weeks after reconstitution, the mice were infected with LCMV and the virus-specific T cells

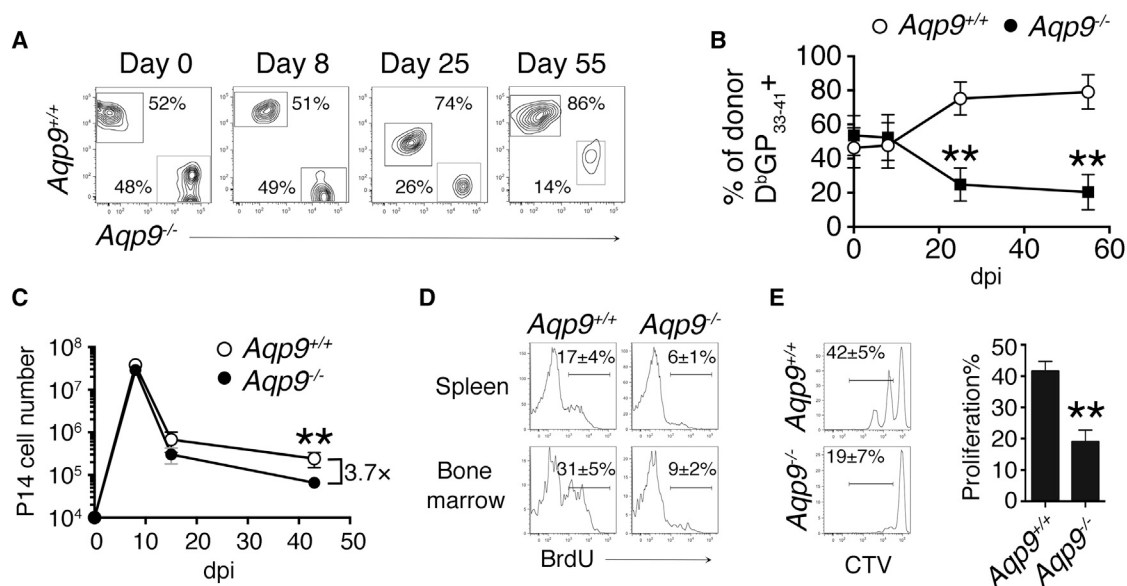


Figure 2. AQP9 Deficiency Impairs Formation of LCMV-Specific Memory CD8⁺ T Cells

(A and B) Bone marrow chimeric mice containing a 1:1 ratio of *Aqp9*^{+/+} (open circles, Ly5.2⁺Thy1.1⁺) and *Aqp9*^{-/-} (black squares, Ly5.2⁺Thy1.2⁺) bone marrow cells were infected with LCMV-Armstrong and the frequency of the two populations within the DGP³³⁻⁴¹-specific CD8⁺ T cells were analyzed longitudinally by flow cytometry.

(C) *Aqp9*^{-/-} or littermate *Aqp9*^{+/+} P14 CD8⁺ T cells (10⁴ cells) were adoptively transferred into B6 mice that were subsequently infected with LCMV-Armstrong. The numbers of donor P14 CD8⁺ T cells were determined at the indicated dpi.

(D) P14 chimeric mice described in (C) were given BrdU drinking water (1 mg/ml) from 30–51 dpi to measure the rates of homeostatic proliferation in the memory CD8⁺ T cells. Amounts of nuclear BrdU were measured by flow cytometry.

(E) *Aqp9*^{-/-} or littermate *Aqp9*^{+/+} P14 CD8⁺ T cells were purified at 25 dpi, labeled with CTV, adoptively transferred to naive B6 mice, and analyzed by flow cytometry for CTV dilution. The bar graph shows the percentages of divided cells.

Data are representative (A, C–E) or cumulative (B) of three and four (C) independent experiments (n = 5–15 mice/group): **p < 0.01 (see also Figure S2).

were analyzed longitudinally. This showed that LCMV-specific *Aqp9*^{-/-} effector CD8⁺ T cell expansion was not affected compared with *Aqp9*^{+/+} cells at 8 dpi. However, the *Aqp9*^{-/-} CD8⁺ T cells revealed a profound defect in their survival thereafter, and the frequency of *Aqp9*^{-/-} memory CD8⁺ T cells steadily declined over time (Figures 2A and 2B). This result demonstrated a critical role for AQP9 in memory CD8⁺ T cell formation and maintenance. To more rigorously examine the CD8⁺ T cell-intrinsic requirement of *Aqp9* in memory CD8⁺ T cell formation, we created P14 mice lacking *Aqp9* and transferred small numbers of naive *Aqp9*^{-/-} or *Aqp9*^{+/+} P14 CD8⁺ T cells into *Aqp9*^{+/+} littermates that were subsequently infected with LCMV. The numbers of donor *Aqp9*^{-/-} or *Aqp9*^{+/+} P14 CD8⁺ T cells were assessed at 8, 15, and 43 dpi (Figure 2C). Similar to the bone marrow chimeras, the P14 CD8⁺ T cells lacking *Aqp9* expanded similarly to their wild-type counterparts, but were poorly maintained during the effector → memory transition and generated a pool of memory CD8⁺ T cells that was ~4-fold smaller than the wild-type (WT) cells (Figure 2C). Furthermore, BrdU-labeling experiments from 30–51 dpi and cell tracer violet (CTV)-labeling experiments revealed that *Aqp9*^{-/-} memory CD8⁺ T cells displayed a profound defect in homeostatic proliferation in the spleen and bone marrow (Figures 2D and 2E). Closer interrogation of the quality of the effector and memory CD8⁺ T cells revealed that the *Aqp9* was required for optimal differentiation of memory CD8⁺ T cells (Figure S1). That is, *Aqp9*^{-/-}

memory CD8⁺ T cells contained fewer IL-7R α ^{hi} CD27^{hi} and CD62L^{hi} cells than the *Aqp9*^{+/+} cells. Thus, there was a block in the development of T_{CM} cells. These studies identified a new protein AQP9 that is critical for memory T cell development and survival after viral infection.

AQP9 Deficiency Impairs Glycerol Uptake and TAG Synthesis in CD8⁺ T Cells

AQP9 transports water, glycerol, and urea. To investigate which solute was involved in AQP9-mediated memory CD8⁺ T cells survival, we attempted to rescue AQP9-deficient CD8⁺ T cells by retroviral (RV) overexpression of *Aqp3* (also permeable to water, glycerol, and urea) or *Aqp1* (permeable to water). This showed that *Aqp3* overexpression could partially rescue *Aqp9*^{-/-} memory CD8⁺ T cell formation, but *Aqp1* could not. This result suggested that glycerol or urea, as opposed to water, were the critical AQP9-dependent solutes for memory T cell formation (Figure S2A). Moreover, as prior studies found aberrantly high levels of glycerol in the serum of *Aqp9*^{-/-} mice because of impaired glycerol uptake by the liver (Rojek et al., 2007), we hypothesized that defective glycerol import in *Aqp9*^{-/-} T cells may contribute to their poor memory T cell survival. In support of this idea, we observed that the *Aqp9*^{-/-} P14 CD8⁺ T cells contained ~50% less intracellular glycerol than their *Aqp9*^{+/+} counterparts following in vitro activation (Figure 3A).

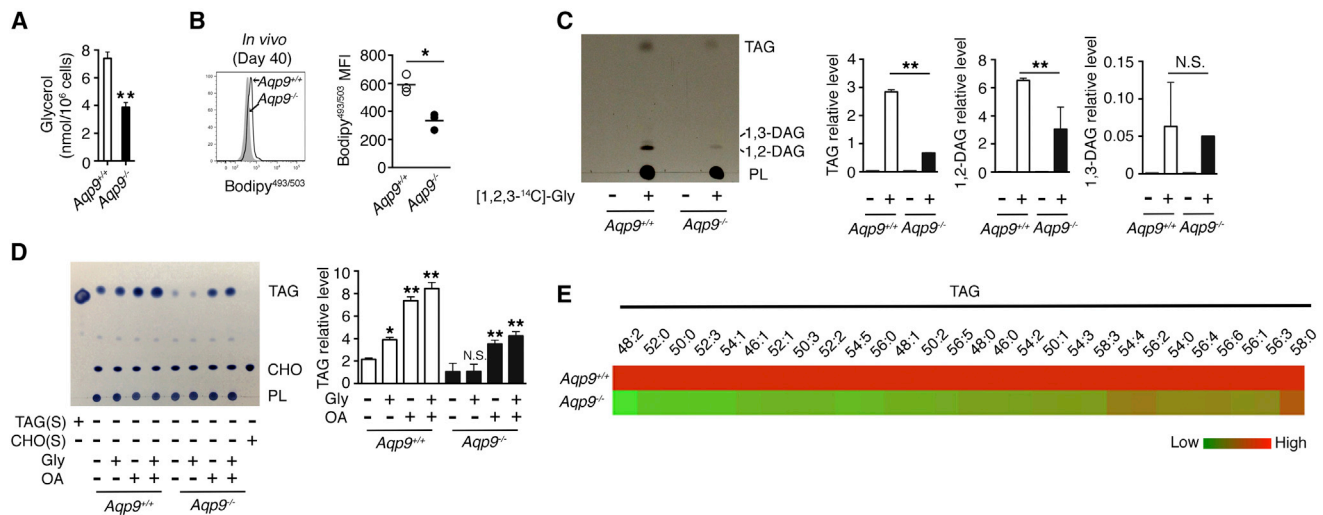


Figure 3. AQP9 Deficiency Impairs TAG Synthesis and Storage in CD8⁺ T Cells

(A) *Aqp9*^{-/-} or *Aqp9*^{+/+} P14 CD8⁺ T cells were cultured in vitro with GP₃₃₋₄₁ peptide for 3 days and then in IL-7 for two days. The amount of free glycerol was measured in total cell lysates by a coupled enzymatic reaction system.

(B) *Aqp9*^{-/-} or *Aqp9*^{+/+} P14 memory CD8 T cells from 40 dpi were stained with the neutral lipid indicator Bodipy^{493/503} and analyzed by flow cytometry.

(C) P14 CD8⁺ T cells described in (A) were pulsed with 0.1 μ Ci/ml [1,2,3-¹⁴C]-Glycerol for 4 hr, then lipids were extracted and resolved by TLC. The bar graphs on the right show densitometry quantification of TAG and DAG autoradiography bands after a 2-week exposure.

(D) P14 CD8⁺ T cells described in (A) were cultured in the presence or absence of glycerol (Gly) or OA for 2 days before lipid extraction and TLC assay. Standards for TAG and CHO were loaded on the left- and right-most lanes. The bar graph on the right shows densitometry quantification of TAG band.

(E) Lipids were extracted from *Aqp9*^{+/+} and *Aqp9*^{-/-} P14 CD8⁺ T cells described in (A) for LC-MS analysis of TAG isobaric species.

Data are representative of two (B and E) and three (A, C, and D) independent experiments (n = 3–7 mice/group): *p < 0.05, **p < 0.01, and not significant (n.s.) (see also Figure S3).

Glycerol is the molecular backbone of TAGs and most phospholipids (PLs). To determine if AQP9 deficiency affected glycerolipid homeostasis in CD8⁺ T cells, we first evaluated the total cellular neutral lipid content (e.g., TAGs) in the *Aqp9*^{-/-} LCMV-specific memory CD8⁺ T cells (40 dpi) using Bodipy^{493/503} labeling and observed that it was approximately one-half that of the *Aqp9*^{+/+} control cells (Figure 3B). The reduction of Bodipy^{493/503} mean fluorescence intensity (MFI) could be due to either decreased TAG synthesis or increased lipolysis or both. To more closely monitor the incorporation of glycerol into TAGs (i.e., synthesis), *Aqp9*^{+/+} or *Aqp9*^{-/-} CD8⁺ T cells were pulsed with radioactive ¹⁴C glycerol (for 4 hr) before lipid extraction and thin layer chromatography (TLC). Glycerol incorporation into diacylglycerol (DAG) and TAG was detected in *Aqp9*^{+/+} cells, but virtually none was detected in *Aqp9*^{-/-} cells (Figure 3C). We then compared the ability of *Aqp9*^{-/-} and *Aqp9*^{+/+} CD8⁺ T cells to synthesize TAGs by culturing the activated CD8⁺ T cells in glycerol and the free FA oleic acid (OA) for 48 hr. This treatment boosted TAG synthesis in *Aqp9*^{+/+} CD8⁺ T cells, but the *Aqp9*^{-/-} cells were considerably less efficient (Figure 3D). Separation of glycerol and OA in the cultures demonstrated that exogenous OA could promote TAG synthesis in both *Aqp9*^{+/+} and *Aqp9*^{-/-} T cells, but glycerol could only enhance TAG synthesis in the *Aqp9*^{+/+} CD8⁺ T cells. This result provided greater evidence that AQP9 was necessary for glycerol import and TAG synthesis in CD8⁺ T cells. To further characterize the various glycerolipid species in the CD8⁺ T cells, we performed lipidomic analysis using liquid chromatog-

raphy-mass spectrometry (LC-MS). This validated the TLC results by demonstrating a marked decrease in all TAG isobaric species in the *Aqp9*^{-/-} CD8⁺ T cells compared with *Aqp9*^{+/+} cells (Figure 3E). Interestingly, the amounts of intracellular PLs or cholesterol (CHO) were only marginally affected by AQP9 deficiency, indicating a more specific defect in TAG biogenesis (Figures 3D and S2B). Altogether, these results show that AQP9 is necessary to maintain normal levels of glycerol and TAGs in antigen-specific CD8⁺ T cells, and when coupled to the data shown in Figure 2, reveal a regulatory mode of TAG synthesis in memory CD8⁺ T cell survival.

AQP9 Deficiency Reduces ATP Levels and Alters Metabolic States in CD8⁺ T Cells

Next, we determined if the reduced amounts of TAGs in the CD8⁺ T cells affected their bioenergetic states. To this end, we compared the rates of glycolysis and mitochondrial respiration using the Seahorse Extracellular Flux Analyzer. This showed that in vitro activated P14 *Aqp9*^{-/-} CD8⁺ T cells had substantially higher extracellular acidification rates (ECAR) (i.e., glycolytic rates) and modestly higher oxygen consumption rates (OCR) (i.e., mitochondrial respiration) than the *Aqp9*^{+/+} cells (Figures 4A and S3). The increased ECAR:OCR ratios in *Aqp9*^{-/-} CD8⁺ T cells indicate a shift toward preferential use of glycolysis over OXPHOS (Figures 4A, right graph, and S3E). Furthermore, *Aqp9* deficiency affected the mitochondrial spare respiratory capacity (SRC) of LCMV-specific CD8 T cells, which is a measurement of the maximal rate of respiration after mitochondrial membrane

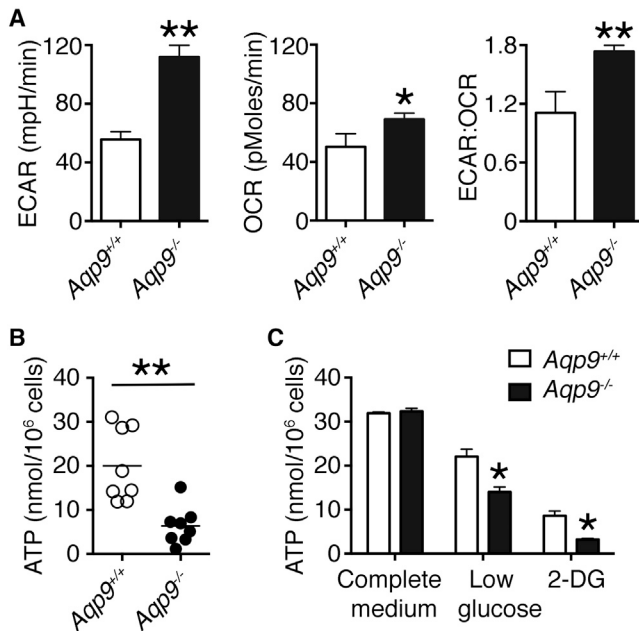


Figure 4. AQP9 Deficiency Reduces ATP Levels and Increases Glycolytic Rates in CD8⁺ T Cells

(A) *Aqp9*^{-/-} or *Aqp9*^{+/+} P14 CD8⁺ T cells were cultured in vitro with GP₃₃₋₄₁ peptide for 3 days and then in IL-7 for two days. Rates of ECAR and OCR were then measured using the Seahorse Extracellular Flux Analyzer. The bar graphs show the basal levels of ECAR, OCR, and the ratio between ECAR and OCR. (B) The amount of intracellular ATP was measured in *Aqp9*^{-/-} or *Aqp9*^{+/+} P14 memory CD8 T cells from 40 dpi by a bioluminescence assay as described in Experimental Procedures.

(C) P14 CD8⁺ T cells described in (A) were cultured in complete medium (11 millimolar [mM] glucose), low glucose medium (2.2 mM glucose), or complete medium plus the glycolysis inhibitor 2-DG and ATP levels were measured 12 hr later.

Data are cumulative from three independent experiments (n = 8 mice/group) (B) or representative of three independent experiments (n = 3 mice/group) (A and C): *p < 0.05 and **p < 0.01.

uncoupling (Figures S3A and S3B). The mitochondrial SRC in memory T cells has been suggested to be affected by FA availability (O'Sullivan et al., 2014), and indeed, *Aqp9*^{-/-} LCMV-immune CD8⁺ T cells had lower amounts of intracellular FFA (Figure S3F). Given that TAGs are an important biofuel that supply FAs, through lipolysis, for mitochondrial FAO, it is likely that TAG insufficiency in *Aqp9*^{-/-} memory CD8⁺ T cells prevents these cells from sustaining high rates of FAO necessary for memory T cell survival.

To determine if the metabolic alterations affected total ATP levels, we measured the amount of intracellular ATP between the two groups of memory P14 CD8⁺ T cells isolated 40 dpi and found indeed, that the *Aqp9*^{-/-} CD8⁺ T cells had markedly lower ATP levels compared with the *Aqp9*^{+/+} cells (Figure 4B). This finding suggested that despite the increase in glycolysis, *Aqp9*^{-/-} CD8⁺ T cells were unable to sustain normal ATP levels. To evaluate this further, in vitro activated *Aqp9*^{-/-} and *Aqp9*^{+/+} P14 CD8⁺ cells were cultured in medium containing low glucose concentrations or 2-deoxy-D-glucose (2-DG), an inhibitor of the first step of glycolysis. This demonstrated that the *Aqp9*^{-/-} CD8⁺ T cells were more sensitive than the *Aqp9*^{+/+} cells to glucose

deprivation based on lowered amounts of ATP (Figure 4C). Together, these findings suggested that AQP9-deficient memory T cells were more reliant on glycolysis, likely because of reduced lipid stores, but nonetheless, were unable to generate sufficient amounts of ATP for long-term survival.

Increased TAG Synthesis Rescues *Aqp9*^{-/-} Memory CD8⁺ T Cell Survival

The above data demonstrated that AQP9 is required for glycerol import to support TAG synthesis and survival in memory CD8⁺ T cells. To further investigate the regulation of TAG synthesis in CD8⁺ T cells during viral infection, we examined the mRNA expression patterns of enzymes involved in TAG synthesis (Figure 5A) in *Aqp9*^{+/+} or *Aqp9*^{-/-} LCMV-specific CD8⁺ T cells as they differentiated from naive → effector → memory CD8⁺ T cells (Rodríguez et al., 2011; Shi and Burn, 2004). This showed that a few of the enzymes were upregulated in the *Aqp9*^{+/+} virus-specific CD8⁺ T cells at early (4.5 dpi) or late (8 dpi) effector time points relative to naive CD8⁺ T cells, but stunningly, all of the enzymes assessed in this process were coordinately upregulated in memory T cells (40 dpi) (Figure 5B). In contrast, several of these enzymes, such as *glycerol kinase* (*Gyk*), *glycerol-3-phosphate acyltransferase mitochondrial* (*Gpat1*), *monoacylglycerol O-acyltransferase 1* (*Mogat1*), and *DAG O-acyltransferase 1* (*Dgat1*), were not elevated in *Aqp9*^{-/-} memory P14 CD8⁺ T cells to the same extent as the WT cells (Figure 5B).

To determine if the reduction of TAG synthesis-related genes in *Aqp9*^{-/-} cells contributed to their defects in memory T cell formation, we transduced *Aqp9*^{-/-} P14 CD8⁺ T cells with RVs overexpressing *Gyk*, *Dgat1*, *Gpat1*, and *Mogat1* or an empty vector (EV) control and analyzed the number of memory CD8 T cells and Bodipy staining at 40 dpi (Figure 5C). In addition, *Aqp9*^{+/+} P14 cells expressing an EV RV were included as a comparison. This showed that overexpression of these genes in the *Aqp9*^{-/-} P14 T cells restored intracellular neutral lipid content and memory CD8 T cell numbers to levels similar to that observed in *Aqp9*^{+/+} T cells (Figure 5C). A longitudinal analysis of these experiments (Figure 5D) revealed that overexpression of the TAG synthesis-related genes had a relatively bigger effect on the frequency of *Aqp9*^{-/-} memory T cells (40 dpi) as opposed to effector cells (8 dpi). Likewise, a similar effect was observed when the genes were overexpressed in *Aqp9*^{+/+} P14 CD8⁺ T cells (Figure 5D). This showed that boosting TAG synthesis had a more prominent effect on the development of memory than on effector CD8 T cells. Notably, *Dgat1* overexpression had a greater impact on *Aqp9*^{-/-} P14 memory CD8⁺ T cell survival than the other enzymes or than that in *Aqp9*^{+/+} cells, suggesting decreased DGAT1 activity is an underlying cause of the *Aqp9*^{-/-} memory CD8 T defect. These results convincingly demonstrate that impaired TAG synthesis and storage is an underlying cause of defective memory CD8⁺ T cell formation in *Aqp9*^{-/-} cells, thereby illuminating the vital role of TAG synthesis in CD8⁺ memory T cell formation and homeostasis.

IL-7 Enhances TAG Synthesis to Promote CD8⁺ Memory T Cell Survival

To better understand how TAG synthesis is regulated in memory CD8⁺ T cells, we returned to IL-7 because our initial data

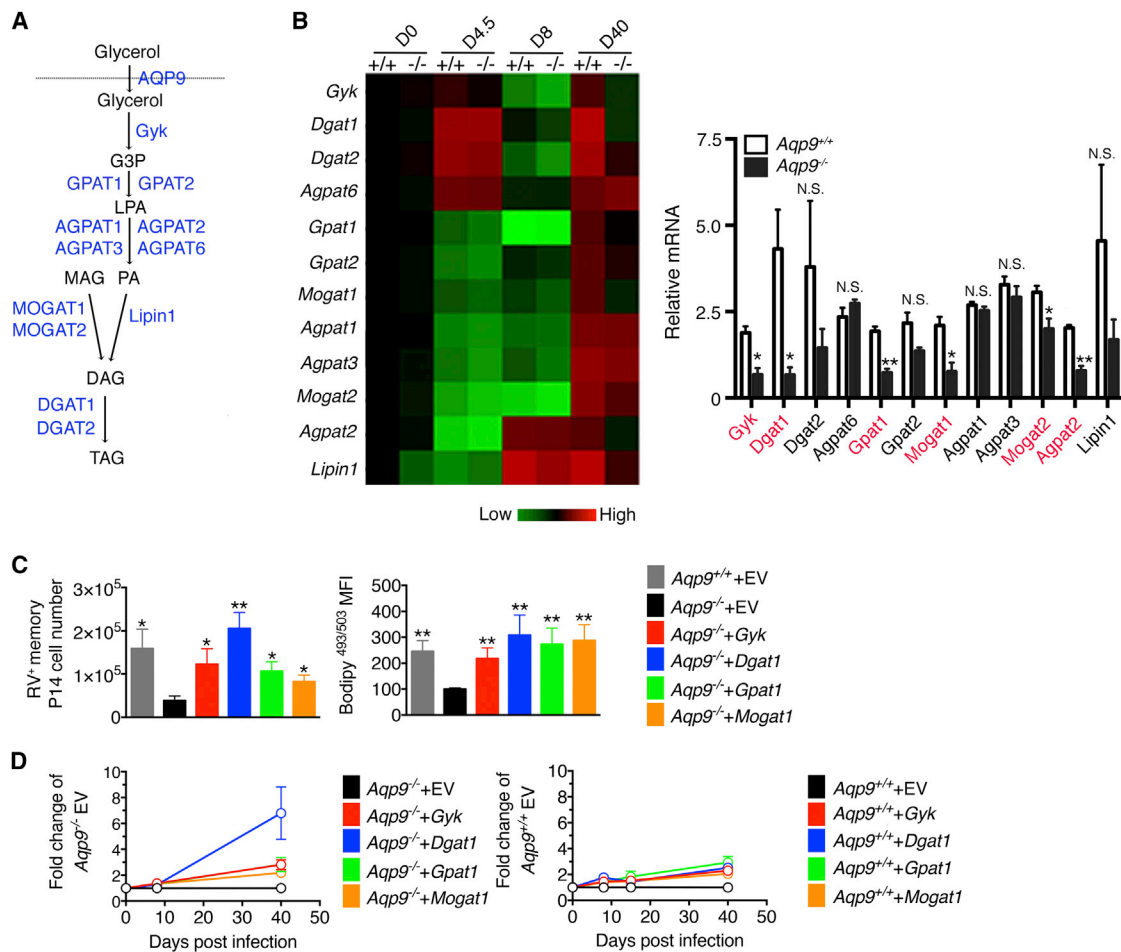


Figure 5. Overexpression of Glycerol Kinase and TAG Synthases Rescues Survival of *Aqp9*^{-/-} Memory CD8⁺ T Cells

(A) Outline of TAG synthesis pathway with AQP9 and TAG synthases highlighted in blue.

(B) Heat map shows the mean mRNA expression level of the indicated genes in virus-specific *Aqp9*^{-/-} or *Aqp9*^{+/+} P14 CD8⁺ T cells purified at 0, 4.5, 8, and 40 dpi using qRT-PCR (values are normalized to naive CD8 T cells [day 0]). The bar graph shows the amount of mRNA relative to L9 mRNA in memory CD8 T cells (40 dpi) and the statistically significant differentially expressed genes are highlighted in red.

(C) *Aqp9*^{-/-} P14 CD8⁺ T cells were transduced with RVs overexpressing *Gyk*, *Dgat1*, *Gpat1*, *Mogat1*, or control EV and adoptively transferred into B6 mice that were subsequently infected with LCMV-Armstrong. *Aqp9*^{+/+} P14 CD8 T cells transduced with EV were included for comparison (dark gray). The bar graphs show the numbers of donor P14 CD8⁺ T cells and MFI of Bodipy^{493/503} staining at 40 dpi.

(D) Longitudinal analysis of the RV-transduced *Aqp9*^{-/-} and *Aqp9*^{+/+} P14 CD8⁺ T cells, as described in (C). At each time point, the frequency of RV⁺ cells was normalized to EV control cells and plotted in the line graph.

Data are cumulative of three (B–D) experiments (n = 6–10 mice/group): *p < 0.05, **p < 0.01, and not significant (n.s.).

indicated that IL-7 was sufficient and necessary for AQP9 expression in antigen-specific CD8⁺ T cells. Given that IL-7R α -signaling is critical for memory CD8⁺ T cell survival (Goldrath et al., 2002; Kaech et al., 2003; Kieper et al., 2002; Prlie et al., 2002; Schluns et al., 2000; Tan et al., 2002), we hypothesized that IL-7 may directly regulate TAG synthesis in these cells. First, we examined if IL-7 signaling was necessary to sustain TAG levels in memory CD8⁺ T cells by transferring such cells into *Il7*^{+/+} or *Il7*^{-/-} animals for 5 days. These experiments revealed that the donor memory CD8⁺ T cells isolated from IL-7-deficient hosts had lower amounts of Bodipy^{493/503} staining relative to the IL-7-sufficient hosts, which indicated that IL-7 was required to sustain neutral lipid levels in memory CD8⁺ T cells (Figure 6A).

Second, stimulating naive or LCMV-specific memory CD8⁺ T cells in vitro with recombinant IL-7 (Figures 6B and 6C) or in vivo with IL-7/anti-IL-7 (M25) mAb complexes (Figure 6D) demonstrated that IL-7 treatment induced lipogenesis and TAG synthesis in memory CD8⁺ T cells profoundly more than in naive CD8⁺ T cells. Additionally, in vitro stimulation with IL-7 induced the expression of several TAG synthesis-related genes in activated, but not naive, CD8 T cells (Figure S4). Together, these findings indicated that, on a per cell basis, memory CD8⁺ T cells had a greater capacity to synthesize neutral lipids (i.e., TAGs) than naive CD8⁺ T cells and suggested that IL-7 induces distinct metabolic programs between naive and memory T cells. Because IL-7 is a critical memory T cell survival factor,

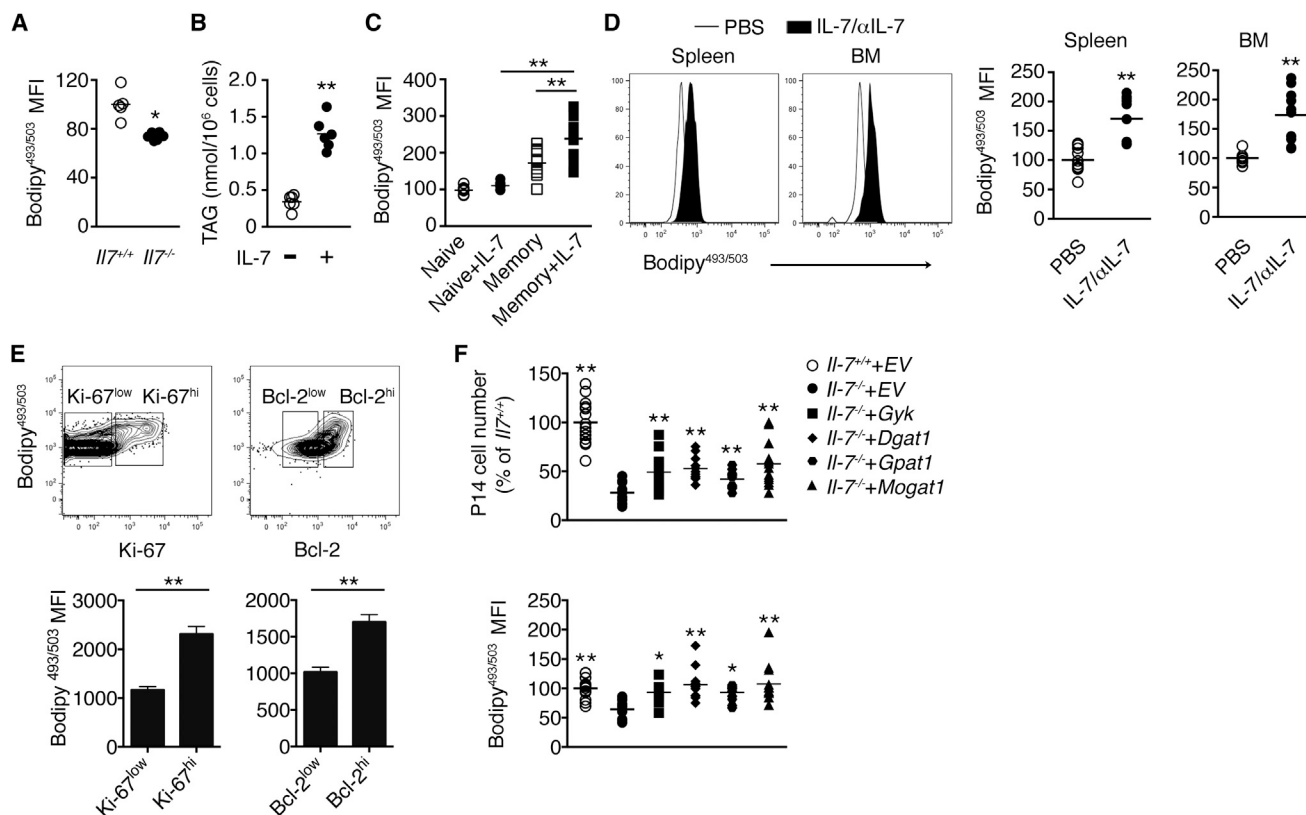


Figure 6. IL-7-Driven Glycerol Metabolism and TAG Synthesis Are Critical for Memory CD8⁺ T Cell Survival

(A and B) Purified memory P14 CD8⁺ T cells from 50–70 dpi were (A) transferred to *IL7*^{+/+} and *IL7*^{-/-} mice and 5 days later the donor cells were stained with Bodipy^{493/503} and analyzed by flow cytometry or (B) stimulated with IL-7 for 12 hr and then TAG levels were measured as described in [Experimental Procedures](#). (C) Purified naive or memory P14 CD8⁺ T cells (from 50–70 dpi) were stimulated with IL-7 for 12 hr and then TAG levels were measured using Bodipy^{493/503} staining and flow cytometry.

(D) P14 chimeric mice containing memory P14 CD8⁺ T cells from 40–60 dpi were injected with IL-7/anti-IL-7 (M25) complex. At 3 days later, TAG levels in donor P14 cells in the spleen and bone marrow were measured using Bodipy^{493/503} staining and flow cytometry.

(E) Memory P14 CD8⁺ T cells purified at 50–150 dpi were stained with Bodipy^{493/503}, Ki-67, and Bcl-2 and analyzed by flow cytometry. The bar graphs show Bodipy^{493/503} MFI in the indicated cell populations.

(F) Memory P14 CD8⁺ T cells (from 30–40 dpi) overexpressing *Gyk*, *Dgat1*, *Gpat1*, *Mogat1*, or EV were adoptively transferred into *IL7*^{-/-} mice. For comparison, the EV-transduced cells were also transferred into *IL7*^{+/+} mice (open circles). At 2–4 weeks later, donor RV⁺ P14 CD8⁺ T cells were enumerated and stained with Bodipy^{493/503}. Values were normalized to those of EV-transduced cells in *IL7*^{+/+} mice (open circles).

Data shown are cumulative of two (D), three (A–C), and four (F) independent experiments or representative of five independent experiments (E) (n = 5–20 mice/group): *p < 0.05 and **p < 0.01 (see also [Figure S4](#)).

we asked if TAG synthesis was associated with memory T cell survival and homeostatic turnover. Interestingly, dividing (Ki-67⁺) memory T cells had higher Bodipy^{493/503} staining than the resting Ki-67⁻ counterparts, suggesting a close connection between lipogenesis and CD8⁺ memory T cell self-renewal ([Figure 6E](#)). Furthermore, the memory CD8⁺ T cells that contained the greatest amount of neutral lipids also had the highest Bcl-2 expression, possibly highlighting a link between TAG levels and memory T cell survival.

Given the strong association between TAG synthesis and memory CD8⁺ T cell homeostatic proliferation and survival, we wondered if increasing TAG synthesis could rescue memory T cell survival in *IL7*^{-/-} hosts. To this end, *Aqp9*^{+/+} P14 memory CD8⁺ T cells overexpressing *Gyk* or TAG synthases (as in [Figure 5D](#)) were transferred into *IL7*^{-/-} mice and analyzed 2–4 weeks

later. This showed that overexpression of these enzymes in memory CD8⁺ T cells could significantly boost neutral lipid levels and partially rescue cell survival in IL-7-deficient hosts ([Figure 6F](#)). These results indicated that TAG synthesis, a process not previously known to be controlled by IL-7, largely contributes to IL-7-mediated memory T cell survival.

IL-7 Drives TAG Synthesis and Promotes Human CD8⁺ T Cell Survival

Finally, we asked if IL-7 signaling could similarly induce *Aqp9* expression and TAG synthesis in human memory T cells. As a first step, we characterized *Aqp9* gene expression in different CD8⁺ T cell subsets, including CCR7⁺CD45RA⁺ naive cells, CCR7⁻CD45RA⁺ T_{EMRA} cells, CCR7⁻CD45RA⁻ T_{EM} cells, and CCR7⁺CD45RA⁻ T_{CM} cells and found that *Aqp9* mRNA was

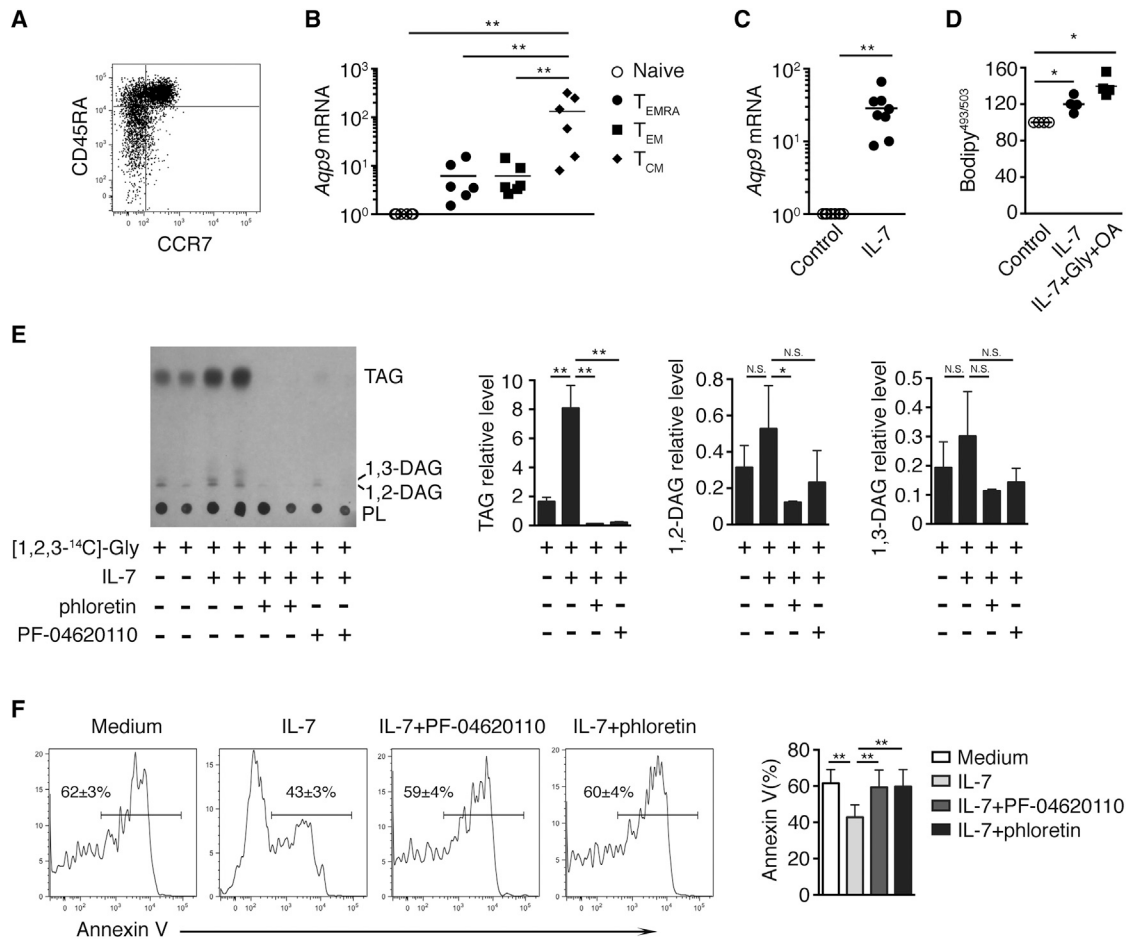


Figure 7. IL-7 Drives TAG Synthesis and Promotes Human Memory CD8⁺ T Cell Survival

(A and B) CCR7⁺ CD45RA⁺ CD8⁺ naive T cells, CCR7⁻ CD45RA⁺ CD8⁺ T_{EMRA}, CCR7⁻ CD45RA⁻ CD8⁺ T_{EM} cells, and CCR7⁺ CD45RA⁻ CD8⁺ T_{CM} cells were FACS purified and *Aqp9* mRNA expression was measured by qRT-PCR (values shown are normalized to naive T cells).

(C) Scatter plot shows *Aqp9* mRNA expression in CD45RA⁻ CD8⁺ T cells stimulated with IL-7 for 24 hr, measured by qRT-PCR, and normalized to the untreated control cells.

(D) Freshly isolated human CD45RA⁻ CD8⁺ T cells were stimulated with or without IL-7 or IL-7 plus glycerol and OA for 24 hr before Bodipy^{493/503} staining. Bodipy^{493/503} MFI was normalized to the untreated control cells.

(E) Freshly isolated human CD45RA⁻ CD8⁺ T cells from two individuals were pulsed with 0.1 μ Ci/ml [1,2,3-¹⁴C]-Glycerol in the absence or presence of IL-7 or the indicated drugs for 4 hr before lipid extraction and TLC assay. The bar graphs on the right show densitometry quantification of TAG and PL autoradiography bands after a 10-week exposure.

(F) Histograms show the percentage of Annexin V⁺ CD45RA⁻ CD8⁺ T cells after a 12 hr treatment as indicated. The bar graph on the right shows the cumulative data of six samples.

Data shown are a cumulative of three (B–D) independent experiments or representative of three (E and F) independent experiments (n = 5–8 subjects/group): *p < 0.05 and **p < 0.01.

expressed to a higher extent in T_{EMRA}, T_{EM}, and T_{CM} cells compared to naive cells, with the T_{CM} cells expressing the most (Figures 7A and 7B). Importantly, IL-7 treatment induced *Aqp9* expression in CD45RA⁻ memory T cells based on qRT-PCR (Figure 7C) and TAG synthesis in human CD8⁺ T cells based on increased Bodipy^{493/503} labeling (Figure 7D). These results suggested the ability of IL-7 to induce AQP9 and promote TAG synthesis was conserved in both human and murine memory T cells.

Next, we cultured human T cells with ¹⁴C-labeled glycerol in the presence or absence of IL-7 to directly measure glycerol up-

take and TAG synthesis by TLC (Figure 7E). Phloretin, an inhibitor of AQPs (Abrami et al., 1996), or PF-04620110, an inhibitor of DGAT1 (Dow et al., 2011) were used as specificity controls. These results showed that IL-7 increased TAG synthesis in human T cells in an AQP9- and DGAT1-dependent manner. Finally, we asked if TAG synthesis contributed to IL-7-mediated CD8⁺ T cell survival in human T cells in vitro. As shown in Figure 7F, the addition of IL-7 to in vitro cultures of CD45RA⁻ memory CD8⁺ T cells augmented T cell survival considerably compared with cultures maintained without IL-7. Moreover, the addition of phloretin or PF-04620110 reversed IL-7-mediated survival

effects, suggesting that IL-7 also promoted survival of human memory T cells in a manner dependent on AQP9 and TAG synthesis.

DISCUSSION

The IL-7/IL-7R signaling axis is a well-established pathway necessary for memory T cell formation and homeostasis (Kaeche et al., 2003; Schluns et al., 2000). The pro-survival effect of IL-7 on lymphocytes has mainly been attributed to the induction of anti-apoptotic factors and enhancement of glut1 expression and glucose metabolism (Barata et al., 2004; Opferman et al., 2003; Rathmell et al., 2001; Schluns et al., 2000; von Freeden-Jeffry et al., 1997; Wofford et al., 2008), but little else is known about how IL-7 controls memory T cell longevity and homeostasis. Our findings illuminate new mechanisms by which IL-7 promotes immunological memory after viral infection through tailoring memory CD8⁺ T cell metabolism and survival via glycerol import (via AQP9) and TAG synthesis and storage.

The importance of FAO in maintaining pathogen-specific memory CD8⁺ T cells after infection has been emphasized by recent studies (van der Windt et al., 2012), however, little is known about how and where memory T cells obtain lipids to sustain FAO long-term. A recent report demonstrated that memory CD8⁺ T cells do not take up free FAs as well as effector CD8⁺ T cells, and the lipids that fuel memory T cell FAO are generated intrinsically through LAL-mediated lipolysis of TAGs (O'Sullivan et al., 2014). This raises the relevant question of what controls TAG availability in memory T cells. Our data shed insight on this question by identifying that IL-7 sustains TAG stores in memory CD8⁺ T cells through the specific induction of glycerol uptake and lipogenesis by AQP9. The ability to store TAGs efficiently may confer upon memory CD8⁺ T cells a greater ability to survive in stressed or nutrient poor niches. For example, *Aqp9*^{-/-} cells produced as much ATP as control cells in the presence of high glucose concentrations, but this was not the case when glucose was limiting. We also found that the *Aqp9*^{-/-} memory CD8 T cells have decreased mRNA expression of several TAG synthases that could further contribute to TAG insufficiency, and indeed, overexpression of *Gyk*, *Gpat1*, *Dgat1*, or *Mogat1* could restore TAG storage and survival in *Aqp9*-deficient memory T cells. However, the *Aqp9*^{-/-} CD8⁺ T cells still contain roughly one-half the normal amount of glycerol, and perhaps, the overexpression of these enzymes boosts TAG synthesis by increasing their competition for the limited glycerol pool. It was interesting that PLs were not as sensitive as TAGs to the decreased glycerol availability in *Aqp9*^{-/-} CD8 T cells. This may be due to differences in the rates of synthesis or half-lives between PLs and TAGs, or possibly, recycling of cytidine diphosphate glucose (CDP)-DAG enables PL synthesis to occur under limited glycerol conditions (Liu et al., 2014).

In addition to IL-7, IL-15 is another critical cytokine that regulates memory T cell homeostasis and self-renewal (Becker et al., 2002; Goldrath et al., 2002; Kennedy et al., 2000; Lodolce et al., 1998; Prlic et al., 2002; Tan et al., 2002). IL-15 can also affect CD8⁺ T cell metabolism and in vitro it has been shown to stimulate CPT1a expression and FAO (van der Windt et al., 2012). Furthermore, IL-15 is known to accelerate lipolysis in adipocytes

in rodents (Barra et al., 2010), and its plasma level is negatively associated with total fat mass (Nielsen et al., 2008). This suggests a possible “store-and-burn” model whereby IL-7 and IL-15 work in concert to trigger both TAG synthesis and lipolysis simultaneously in memory CD8⁺ T cells to sustain lipid supplies and FAO. However, further in vivo studies are needed to more precisely define the relationship between IL-7 and IL-15 signaling on TAG metabolism in memory CD8⁺ T cells and the metabolic regulation of memory T cell survival and self-renewal.

Both naive and memory T cells express the IL-7R and rely on it for survival (Kaeche et al., 2003; Schluns et al., 2000; Tan et al., 2002; von Freeden-Jeffry et al., 1995). Therefore, another important question is how naive and memory T cells avoid competition for IL-7. Our data indicate that one answer to this question is that naive and memory T cells have adopted different metabolic responses downstream of IL-7R that help diversify the kinds of nutrients utilized by the two types of T cells. IL-7 promotes glucose utilization in T cells (Barata et al., 2004; Wofford et al., 2008), but its added ability to preferentially induce AQP9 and TAG synthesis in memory CD8⁺ T cells endows memory cells with another capability to utilize glycerol and lipids more effectively than naive CD8⁺ T cells. The biochemical basis for the differential responses of memory and naive T cells to IL-7 is not clear, but this provides a mechanism by which naive and memory T cells may avoid nutrient competition, thereby maximizing both T cell numbers and diversity in the periphery. However, it is important to emphasize that the differential dependence of naive and memory T cells on other critical survival factors such as self-major histocompatibility complex (MHC)/peptide and IL-15, respectively (Ge et al., 2004; Murali-Krishna et al., 1999; Swain et al., 1999; Tan et al., 2002; Tanchot et al., 1997), are equally important mechanisms that prevent resource competition between naive and memory T cells and regulate their homeostasis.

Another interesting observation in this study is the association between TAG synthesis and memory T cell homeostatic turnover. Proliferating cells have higher lipid content compared with their non-proliferating counterparts. Our work did not distinguish whether lipid storage fuels cell proliferation or homeostatic proliferation drives TAG synthesis, however, it is possible that the synthesis of TAGs that occurs as memory T cells divide is important for refilling their “gas tank” to sustain FAO and ATP generation during quiescent, non-cycling stages, or under growth factor- or nutrient-poor conditions. In this model, niches rich in IL-7 (and IL-15) could be considered as memory T cell gas stations for bioenergetic refueling. How this concept relates to the maintenance and longevity of other types of adult tissue stem cells will be of great interest. This study highlights that regulation of glycerol stores is critical for memory T cell development, but it also illuminates a previously unrecognized feature that T cells utilize different mechanisms to regulate their glycerol stores at different stages of immune responses. For instance, neither AQP9 nor IL-7R signaling is required during effector CD8⁺ T cell clonal expansion (Schluns et al., 2000), which is associated with a burst of lipogenesis (O'Sullivan et al., 2014 and data not shown). Therefore, the glycerol must be obtained from alternative (AQP9-independent) pathways in effector T cells, possibly through increased abundance of glycolytic intermediates (e.g., dihydroxyacetone phosphate [DHAP]) that

can be converted to glycerol. However, as the effector T cells transition to memory cells and the rates of aerobic glycolysis decline, our work demonstrates that the T cells adopt a new method to obtain sufficient amounts of glycerol. Since naive T cells do not express AQP9, it will be interesting to learn how they regulate glycerol stores.

In summary, this work identifies a previously uncharacterized role of IL-7 in regulating lipid metabolism and maintaining cellular bioenergetics in memory CD8⁺ T cells, which have important implications for development of vaccines and immunotherapies, especially those involving IL-7. Previous studies have demonstrated that recombinant IL-7 is a promising vaccine adjuvant or therapeutic option for treatment of chronic infection or cancer. For example, in mice, IL-7 treatment could boost antiviral immunity following vaccination or lead to better control of chronic LCMV infection (Nanjappa et al., 2008; Pellegrini et al., 2011). In humans, it could increase circulating CD4⁺ and CD8⁺ T cell numbers and homeostatic proliferation in patients infected with HIV, HBV, HCV, or those that have melanoma, renal cell carcinoma, and other cancers (Rosenberg et al., 2006; Sereti et al., 2009; Sportes et al., 2010). Given our discoveries on IL-7 regulating TAG pools in T cells, it will be important to consider how this contributes to the benefits of IL-7-based therapies and whether IL-7 has similar effects on non-immune cells. Additionally, our work uncovers possibilities for boosting or inhibiting T cell memory through drugs that manipulate rates of TAG synthesis.

EXPERIMENTAL PROCEDURES

Mice and Infections

Aqp9^{-/-} mice were kindly provided by Drs. Aleksandra Rojek and Soren Nielsen in Aarhus University (Aarhus, Denmark) via Dr. Peter Agre in Johns Hopkins University (Rojek et al., 2007). *IL7*^{-/-} mice were provided by Schering-Plough Biopharma (Palo Alto, CA). Mice were infected with LCMV-Armstrong (intraperitoneally [i.p.] with 2×10^5 plaque-forming unit [pfu]). All the studies have been approved by Yale University Institutional Animal Care and Use Committee.

TLC

Lipids were loaded onto TLC plates and resolved in heptane/isopropyl ether/acetic acid (60:40:4) solution. The TLC plates were developed in cerium molybdate solution and lipid classifications were made using known lipid standards. For ¹⁴C-glycerol labeling studies, the signals were detected by autoradiography.

LC-MS

Aqp9^{-/-} or littermate *Aqp9*^{+/+} P14 CD8⁺ cells were activated with GP₃₃₋₄₁ peptide for 3 days and then stimulated with IL-7 for another 2 days. At least 1×10^7 cells were used for lipid extraction by methanol and CH₂Cl₂. After drying down in an argon evaporator, lipids were dissolved and loaded for LC-MS assay by LIPID MAPS Lipidomics Core at the University of California, San Diego, CA.

Metabolic Assays

In vitro activated *Aqp9*^{-/-} or littermate *Aqp9*^{+/+} P14 CD8⁺ cells were stimulated with IL-7 for 2 days. After that, cells were washed and plated in XF media containing glucose, L-glutamine, and sodium pyruvate. OCR and ECAR were measured using the XF-96 Extracellular Flux Analyzer (Seahorse Bioscience).

Measurement of ATP, Glycerol, and TAG

Details of these procedures are found in Supplemental Information.

Statistical Analysis

Where indicated, p values were determined by a two-tailed Student's t test. $p < 0.05$ was considered statistically significant. All the data were presented as mean \pm SEM (error bar). In scatter plots, a short black line represented the mean.

For more information, see Extended Experimental Procedures.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures and four figures and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.03.021>.

AUTHOR CONTRIBUTIONS

G.C., M.M.S., S.M.G., P.-C.H., R.A.A., J.W., and S.M.K. designed the experiments. G.C., M.M.S., and S.M.G. performed the experiments. G.C. and S.M.K. wrote the manuscript.

ACKNOWLEDGMENTS

We thank Drs. Aleksandra Rojek, Soren Nielsen, and Peter Agre for providing *Aqp9*^{-/-} mice. We thank Drs. Gerald Shulman and Rachel Perry for their discussion. We thank Barbara Siconolfi for collecting human samples. Drs. Huajin Wang and Tobias Walther provided generous help with TLC assay. This study is supported by grants from the NIH R37AI066232/R01AI074699 (S.M.K.), T32GM07205 (S.M.G.), U19AI089992 (D. Hafler, Yale University), Howard Hughes Medical Institute (S.M.K.), and a Trudeau Fellowship (G.C.).

Received: September 16, 2014

Revised: December 27, 2014

Accepted: February 23, 2015

Published: May 7, 2015

REFERENCES

- Abrami, L., Berthoud, V., Deen, P.M., Rousset, G., Tacnet, F., and Ri-poche, P. (1996). Glycerol permeability of mutant aquaporin 1 and other AQP-MIP proteins: inhibition studies. *Pflügers Arch.* 431, 408–414.
- Akashi, K., Kondo, M., von Freeden-Jeffry, U., Murray, R., and Weissman, I.L. (1997). Bcl-2 rescues T lymphopoiesis in interleukin-7 receptor-deficient mice. *Cell* 89, 1033–1041.
- Barata, J.T., Silva, A., Brandao, J.G., Nadler, L.M., Cardoso, A.A., and Boussiotis, V.A. (2004). Activation of PI3K is indispensable for interleukin 7-mediated viability, proliferation, glucose use, and growth of T cell acute lymphoblastic leukemia cells. *J. Exp. Med.* 200, 659–669.
- Barra, N.G., Reid, S., MacKenzie, R., Werstuck, G., Trigatti, B.L., Richards, C., Holloway, A.C., and Ashkar, A.A. (2010). Interleukin-15 contributes to the regulation of murine adipose tissue and human adipocytes. *Obesity (Silver Spring)* 18, 1601–1607.
- Becker, T.C., Wherry, E.J., Boone, D., Murali-Krishna, K., Antia, R., Ma, A., and Ahmed, R. (2002). Interleukin 15 is required for proliferative renewal of virus-specific memory CD8 T cells. *J. Exp. Med.* 195, 1541–1548.
- Best, J.A., Blair, D.A., Knell, J., Yang, E., Mayya, V., Doedens, A., Dustin, M.L., and Goldrath, A.W.; Immunological Genome Project Consortium (2013). Transcriptional insights into the CD8(+) T cell response to infection and memory T cell formation. *Nat. Immunol.* 14, 404–412.
- Beura, L.K., and Masopust, D. (2014). SnapShot: resident memory T cells. *Cell* 157, 1488–1488 e1481.
- Carbrey, J.M., Gorelick-Feldman, D.A., Kozono, D., Praetorius, J., Nielsen, S., and Agre, P. (2003). Aquaglyceroporin AQP9: solute permeation and metabolic control of expression in liver. *Proc. Natl. Acad. Sci. USA* 100, 2945–2950.
- Dow, R.L., Li, J.C., Pence, M.P., Gibbs, E.M., LaPerle, J.L., Litchfield, J., Piotrowski, D.W., Munchhof, M.J., Manion, T.B., Zavadoski, W.J., et al. (2011).

Discovery of PF-04620110, a potent, selective, and orally bioavailable inhibitor of DGAT-1. *ACS Med. Chem. Lett.* 2, 407–412.

Fox, C.J., Hammerman, P.S., and Thompson, C.B. (2005). Fuel feeds function: energy metabolism and the T-cell response. *Nat. Rev. Immunol.* 5, 844–852.

Ge, Q., Bai, A., Jones, B., Eisen, H.N., and Chen, J. (2004). Competition for self-peptide-MHC complexes and cytokines between naive and memory CD8⁺ T cells expressing the same or different T cell receptors. *Proc. Natl. Acad. Sci. USA* 101, 3041–3046.

Goldrath, A.W., Sivakumar, P.V., Glaccum, M., Kennedy, M.K., Bevan, M.J., Benoist, C., Mathis, D., and Butz, E.A. (2002). Cytokine requirements for acute and Basal homeostatic proliferation of naive and memory CD8⁺ T cells. *J. Exp. Med.* 195, 1515–1522.

Joshi, N.S., Cui, W., Chande, A., Lee, H.K., Urso, D.R., Hagman, J., Gapin, L., and Kaech, S.M. (2007). Inflammation directs memory precursor and short-lived effector CD8⁺ T cell fates via the graded expression of T-bet transcription factor. *Immunity* 27, 281–295.

Kaech, S.M., Tan, J.T., Wherry, E.J., Konieczny, B.T., Surh, C.D., and Ahmed, R. (2003). Selective expression of the interleukin 7 receptor identifies effector CD8 T cells that give rise to long-lived memory cells. *Nat. Immunol.* 4, 1191–1198.

Kennedy, M.K., Glaccum, M., Brown, S.N., Butz, E.A., Viney, J.L., Embers, M., Matsuki, N., Charrier, K., Sedger, L., Willis, C.R., et al. (2000). Reversible defects in natural killer and memory CD8 T cell lineages in interleukin 15-deficient mice. *J. Exp. Med.* 191, 771–780.

Khaled, A.R., Li, W.Q., Huang, J., Fry, T.J., Khaled, A.S., Mackall, C.L., Muegge, K., Young, H.A., and Durum, S.K. (2002). Bax deficiency partially corrects interleukin-7 receptor alpha deficiency. *Immunity* 17, 561–573.

Kieper, W.C., Tan, J.T., Bondi-Boyd, B., Gapin, L., Sprent, J., Ceredig, R., and Surh, C.D. (2002). Overexpression of interleukin (IL)-7 leads to IL-15-independent generation of memory phenotype CD8⁺ T cells. *J. Exp. Med.* 195, 1533–1539.

Kondrack, R.M., Harbertson, J., Tan, J.T., McBreen, M.E., Surh, C.D., and Bradley, L.M. (2003). Interleukin 7 regulates the survival and generation of memory CD4 cells. *J. Exp. Med.* 198, 1797–1806.

Lass, A., Zimmermann, R., Oberer, M., and Zechner, R. (2011). Lipolysis - a highly regulated multi-enzyme complex mediates the catabolism of cellular fat stores. *Prog. Lipid Res.* 50, 14–27.

Lenz, D.C., Kurz, S.K., Lemmens, E., Schoenberger, S.P., Sprent, J., Oldstone, M.B., and Homann, D. (2004). IL-7 regulates basal homeostatic proliferation of antiviral CD4⁺ T cell memory. *Proc. Natl. Acad. Sci. USA* 101, 9357–9362.

Liu, X., Yin, Y., Wu, J., and Liu, Z. (2014). Structure and mechanism of an intramembrane liponucleotide synthetase central for phospholipid biosynthesis. *Nat. Commun.* 5, 4244.

Lodolce, J.P., Boone, D.L., Chai, S., Swain, R.E., Dassopoulos, T., Trettin, S., and Ma, A. (1998). IL-15 receptor maintains lymphoid homeostasis by supporting lymphocyte homing and proliferation. *Immunity* 9, 669–676.

Maraskovsky, E., O'Reilly, L.A., Teepe, M., Corcoran, L.M., Peschon, J.J., and Strasser, A. (1997). Bcl-2 can rescue T lymphocyte development in interleukin-7 receptor-deficient mice but not in mutant rag-1^{-/-} mice. *Cell* 89, 1011–1019.

Murali-Krishna, K., Lau, L.L., Sambhara, S., Lemonnier, F., Altman, J., and Ahmed, R. (1999). Persistence of memory CD8 T cells in MHC class I-deficient mice. *Science* 286, 1377–1381.

Nanjappa, S.G., Walent, J.H., Morre, M., and Suresh, M. (2008). Effects of IL-7 on memory CD8 T cell homeostasis are influenced by the timing of therapy in mice. *J. Clin. Invest.* 118, 1027–1039.

Nielsen, A.R., Hojman, P., Erikstrup, C., Fischer, C.P., Plomgaard, P., Mounier, R., Mortensen, O.H., Broholm, C., Taudorf, S., Krogh-Madsen, R., et al. (2008). Association between interleukin-15 and obesity: interleukin-15 as a potential regulator of fat mass. *J. Clin. Endocrinol. Metab.* 93, 4486–4493.

O'Sullivan, D., van der Windt, G.J., Huang, S.C., Curtis, J.D., Chang, C.H., Buck, M.D., Qiu, J., Smith, A.M., Lam, W.Y., DiPlato, L.M., et al. (2014). Mem-

ory CD8⁺ T cells use cell-intrinsic lipolysis to support the metabolic programming necessary for development. *Immunity* 41, 75–88.

Opferman, J.T., Letai, A., Beard, C., Sorcinelli, M.D., Ong, C.C., and Korsmeyer, S.J. (2003). Development and maintenance of B and T lymphocytes requires antiapoptotic MCL-1. *Nature* 426, 671–676.

Pearce, E.L., Walsh, M.C., Cejas, P.J., Harms, G.M., Shen, H., Wang, L.S., Jones, R.G., and Choi, Y. (2009). Enhancing CD8 T-cell memory by modulating fatty acid metabolism. *Nature* 460, 103–107.

Pearson, C., Silva, A., and Seddon, B. (2012). Exogenous amino acids are essential for interleukin-7 induced CD8 T cell growth. [corrected]. *PLoS ONE* 7, e33998.

Pellegrini, M., Bouillet, P., Robati, M., Belz, G.T., Davey, G.M., and Strasser, A. (2004). Loss of Bim increases T cell production and function in interleukin 7 receptor-deficient mice. *J. Exp. Med.* 200, 1189–1195.

Pellegrini, M., Calzascia, T., Toe, J.G., Preston, S.P., Lin, A.E., Elford, A.R., Shahinian, A., Lang, P.A., Lang, K.S., Morre, M., et al. (2011). IL-7 engages multiple mechanisms to overcome chronic viral infection and limit organ pathology. *Cell* 144, 601–613.

Peschon, J.J., Morrissey, P.J., Grabstein, K.H., Ramsdell, F.J., Maraskovsky, E., Gliniak, B.C., Park, L.S., Ziegler, S.F., Williams, D.E., Ware, C.B., et al. (1994). Early lymphocyte expansion is severely impaired in interleukin 7 receptor-deficient mice. *J. Exp. Med.* 180, 1955–1960.

Prlic, M., Lefrançois, L., and Jameson, S.C. (2002). Multiple choices: regulation of memory CD8 T cell generation and homeostasis by interleukin (IL)-7 and IL-15. *J. Exp. Med.* 195, F49–F52.

Rathmell, J.C., Farkash, E.A., Gao, W., and Thompson, C.B. (2001). IL-7 enhances the survival and maintains the size of naive T cells. *J. Immunol.* 167, 6869–6876.

Rodríguez, A., Catalán, V., Gómez-Ambrosi, J., and Frühbeck, G. (2011). Aquaglyceroporins serve as metabolic gateways in adiposity and insulin resistance control. *Cell Cycle* 10, 1548–1556.

Rojek, A.M., Skowronski, M.T., Füchtbauer, E.M., Füchtbauer, A.C., Fenton, R.A., Agre, P., Frøkiaer, J., and Nielsen, S. (2007). Defective glycerol metabolism in aquaporin 9 (AQP9) knockout mice. *Proc. Natl. Acad. Sci. USA* 104, 3609–3614.

Rosenberg, S.A., Sportès, C., Ahmadzadeh, M., Fry, T.J., Ngo, L.T., Schwarz, S.L., Stetler-Stevenson, M., Morton, K.E., Mavroukakis, S.A., Morre, M., et al. (2006). IL-7 administration to humans leads to expansion of CD8⁺ and CD4⁺ cells but a relative decrease of CD4⁺ T-regulatory cells. *J. Immunother.* 29, 313–319.

Schluns, K.S., Kieper, W.C., Jameson, S.C., and Lefrançois, L. (2000). Interleukin-7 mediates the homeostasis of naive and memory CD8 T cells in vivo. *Nat. Immunol.* 1, 426–432.

Sereti, I., Dunham, R.M., Spritzler, J., Aga, E., Proschan, M.A., Medvik, K., Battaglia, C.A., Landay, A.L., Pahwa, S., Fischl, M.A., et al.; ACTG 5214 Study Team (2009). IL-7 administration drives T cell-cycle entry and expansion in HIV-1 infection. *Blood* 113, 6304–6314.

Shi, Y., and Burn, P. (2004). Lipid metabolic enzymes: emerging drug targets for the treatment of obesity. *Nat. Rev. Drug Discov.* 3, 695–710.

Sportes, C., Babb, R.R., Krumlauf, M.C., Hakim, F.T., Steinberg, S.M., Chow, C.K., Brown, M.R., Fleisher, T.A., Noel, P., Maric, I., et al. (2010). Phase I study of recombinant human interleukin-7 administration in subjects with refractory malignancy. *Clin. Cancer Res.* 16, 727–735.

Swain, S.L., Hu, H., and Huston, G. (1999). Class II-independent generation of CD4 memory T cells from effectors. *Science* 286, 1381–1383.

Tan, J.T., Ernst, B., Kieper, W.C., LeRoy, E., Sprent, J., and Surh, C.D. (2002). Interleukin (IL)-15 and IL-7 jointly regulate homeostatic proliferation of memory phenotype CD8⁺ cells but are not required for memory phenotype CD4⁺ cells. *J. Exp. Med.* 195, 1523–1532.

Tanchot, C., Lemonnier, F.A., Pérarnau, B., Freitas, A.A., and Rocha, B. (1997). Differential requirements for survival and proliferation of CD8 naive or memory T cells. *Science* 276, 2057–2062.

van der Windt, G.J., Everts, B., Chang, C.H., Curtis, J.D., Freitas, T.C., Amiel, E., Pearce, E.J., and Pearce, E.L. (2012). Mitochondrial respiratory capacity is a critical regulator of CD8⁺ T cell memory development. *Immunity* 36, 68–78.

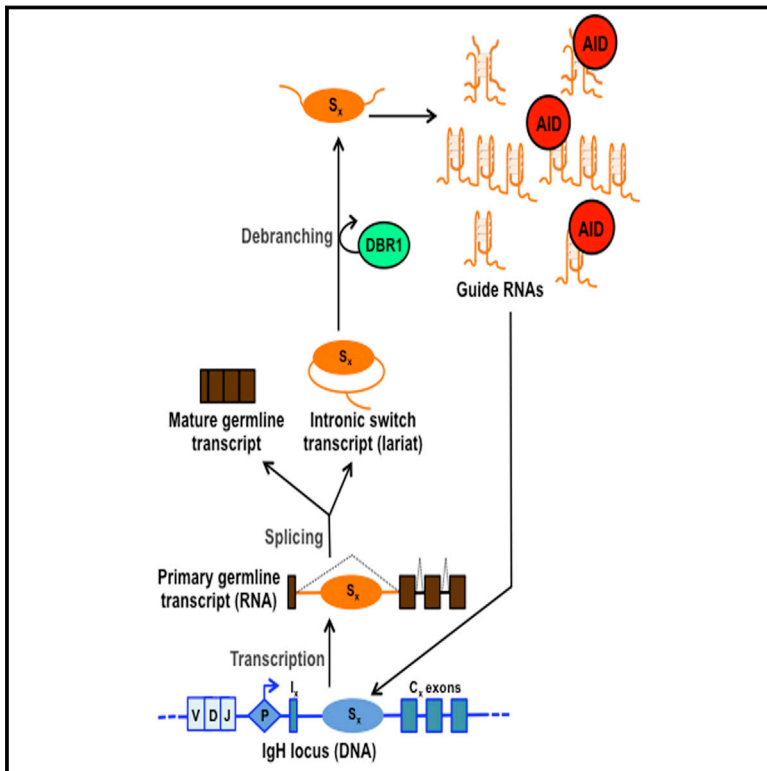
von Freeden-Jeffry, U., Vieira, P., Lucian, L.A., McNeil, T., Burdach, S.E., and Murray, R. (1995). Lymphopenia in interleukin (IL)-7 gene-deleted mice identifies IL-7 as a nonredundant cytokine. *J. Exp. Med.* 181, 1519–1526.

von Freeden-Jeffry, U., Solvason, N., Howard, M., and Murray, R. (1997). The earliest T lineage-committed cells depend on IL-7 for Bcl-2 expression and normal cell cycle progression. *Immunity* 7, 147–154.

Wofford, J.A., Wieman, H.L., Jacobs, S.R., Zhao, Y., and Rathmell, J.C. (2008). IL-7 promotes Glut1 trafficking and glucose uptake via STAT5-mediated activation of Akt to support T-cell survival. *Blood* 111, 2101–2111.

Non-coding RNA Generated following Lariat Debranching Mediates Targeting of AID to DNA

Graphical Abstract



Authors

Simin Zheng, Bao Q. Vuong, ..., Feng-Ting Huang, Jayanta Chaudhuri

Correspondence

chaudhuj@mskcc.org

In Brief

Transcription through the immunoglobulin switch (S) region produces a non-coding RNA that guides the enzyme AID to DNA in a sequence-specific manner to promote antibody class switch recombination.

Highlights

- Switch RNAs can fold into G-quadruplex structures
- AID binds directly to G-quadruplex structures formed by switch RNAs
- Switch RNAs act as AID guides after lariat processing
- Switch RNAs target AID to DNA in a sequence-specific manner



Non-coding RNA Generated following Lariat Debranching Mediates Targeting of AID to DNA

Simin Zheng,^{1,2} Bao Q. Vuong,^{1,3} Bharat Vaidyanathan,^{1,2} Jia-Yu Lin,⁴ Feng-Ting Huang,⁴ and Jayanta Chaudhuri^{1,2,*}

¹Immunology Program, Memorial Sloan Kettering Cancer Center, New York, NY 10065, USA

²Immunology and Microbial Pathogenesis Program, Weill Cornell Graduate School of Medical Sciences, New York, NY 10065, USA

³Department of Biology, City College of New York, New York, NY 10031, USA

⁴Department of Biochemical Science and Technology, College of Life Science, National Taiwan University, Taipei 106, Taiwan

*Correspondence: chaudhuri@mskcc.org

<http://dx.doi.org/10.1016/j.cell.2015.03.020>

SUMMARY

Transcription through immunoglobulin switch (S) regions is essential for class switch recombination (CSR), but no molecular function of the transcripts has been described. Likewise, recruitment of activation-induced cytidine deaminase (AID) to S regions is critical for CSR; however, the underlying mechanism has not been fully elucidated. Here, we demonstrate that intronic switch RNA acts in *trans* to target AID to S region DNA. AID binds directly to switch RNA through G-quadruplexes formed by the RNA molecules. Disruption of this interaction by mutation of a key residue in the putative RNA-binding domain of AID impairs recruitment of AID to S region DNA, thereby abolishing CSR. Additionally, inhibition of RNA lariat processing leads to loss of AID localization to S regions and compromises CSR; both defects can be rescued by exogenous expression of switch transcripts in a sequence-specific manner. These studies uncover an RNA-mediated mechanism of targeting AID to DNA.

INTRODUCTION

Following antigen receptor assembly, mature B cells home to peripheral lymphoid organs where they encounter antigens and undergo immunoglobulin (Ig) heavy-chain (*Igh*) class switch recombination (CSR). CSR is a deletional-recombination event that exchanges the default C_μ constant region gene (C_H) for one of several downstream C_H segments (C_γ, C_ε, or C_α). The reaction proceeds through the introduction of DNA double-strand breaks (DSBs) into transcribed, repetitive DNA elements, called switch (S) regions that precede each C_H gene segment. End joining of DSBs between a donor (S_μ) and a downstream acceptor S region deletes the intervening DNA and juxtaposes a new C_H gene to the variable region gene segment. The B cell thereby “switches” from expressing IgM to one producing IgG, IgE, or IgA, with each secondary isotype having a distinct

effector function during an immune response (Matthews et al., 2014).

The single-strand DNA-specific cytidine deaminase activation-induced cytidine deaminase (AID) is essential for CSR (Muramatsu et al., 2000; Revy et al., 2000). AID deaminates cytosines within transcribed S regions (Chaudhuri et al., 2003; Maul et al., 2011), and the deaminated DNA engages the ubiquitous base-excision and mismatch repair machineries to generate DSBs that are required for CSR (Petersen-Mahrt et al., 2002). A failure to efficiently recruit AID to S regions impairs CSR (Nowak et al., 2011; Pavri et al., 2010; Xu et al., 2010). Conversely, mistargeting of AID activity to non-Ig genes has been implicated in chromosomal translocations and pathogenesis of B cell lymphomas (Nussenzweig and Nussenzweig, 2010; Pasqualucci et al., 2008). While AID is phosphorylated at multiple residues, including at serine-38, phosphorylation is not required for DNA binding (Matthews et al., 2014).

Thus, the molecular mechanisms by which AID is specifically targeted to S regions continue to be an active area of investigation. Transcription through S regions is essential for CSR and is closely linked to the mechanism by which AID specifically binds and gains access to S regions during CSR (Matthews et al., 2014). Each of the C_H genes is organized as individual transcription units comprising of a cytokine inducible promoter, an intervening I-exon, S region, and C_H exons. Splicing of the primary transcript joins the I- and C_H exons to generate a non-coding mature transcript and releases the intronic switch sequence. Transcription through S regions, 1- to 12-kb-long repetitive DNA elements with a guanine-rich non-template strand, predisposes formation of RNA:DNA hybrid structures such as R-loops that expose single-stranded DNA substrates for AID (Matthews et al., 2014). Germline transcription is also required for the binding of AID at S regions through the ability of AID to interact with components of RNA polymerase II (Nambu et al., 2003; Pavri et al., 2010). Both R-loop formation and RNA polymerase-II-mediated recruitment of AID relies on the process of transcription, but the role of germline switch transcripts themselves in the recombination reaction has yet to be identified.

Several intriguing reports have suggested that germline switch transcripts might have mechanistic roles in CSR. Deletion of the Iγ1 exon splice donor site, which inhibits splicing of the primary

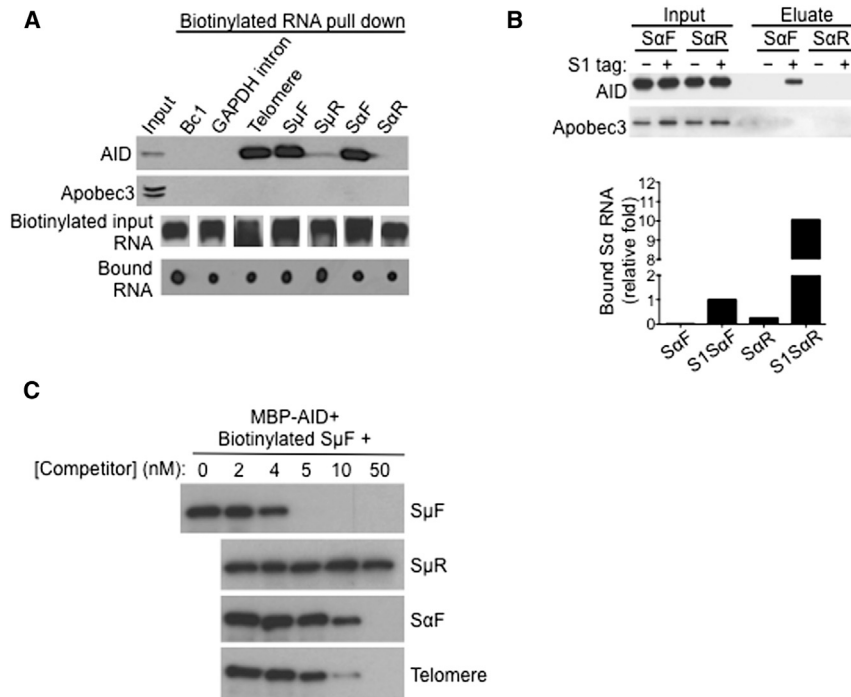


Figure 1. Switch RNA Can Bind AID

(A) In-vitro-transcribed (IVT) telomeric and switch RNAs bind to AID. IVT biotinylated RNAs were folded and incubated with whole-cell extracts from stimulated CH12 cells, followed by pull-down with streptavidin beads. Proteins recovered were analyzed by immunoblot with AID or Apobec3 antibodies, while bound RNAs were analyzed by dot blot using streptavidin-HRP. Input RNAs were verified to be biotinylated by streptavidin-HRP-northern blot. S μ F, S α F, forward/sense switch μ and α RNA; S μ R, S α R; reverse/anti-sense switch μ and α RNA. Result shown is representative of three independent pull-down experiments.

(B) Switch RNA interacts with AID in vivo. CH12 cells stably expressing S1-expressing S α transcripts, in either the forward/sense (S1S α F) or the reverse/anti-sense (S1S α R) orientation, were stimulated. Untagged S α F- and S α R-expressing cells were used as controls. The S1-apptamer tag has affinity for streptavidin and ribonucleoprotein complexes were isolated on streptavidin beads. RNA in the eluates was reverse transcribed and analyzed by qPCR (RT-qPCR) for amounts of S α transcripts relative to S1S α F, while proteins in the eluates were analyzed by immunoblot. Result shown is representative of three independent pull-down experiments.

(C) Competition RNA-binding assay. RNA pull-

down was performed with 1 nM biotinylated S μ F RNA and 100 ng MBP-AID(WT) protein, in the presence of increasing concentrations of non-biotinylated competitor RNAs. Bound MBP-AID(WT) recovered by pull-down with streptavidin beads were analyzed by immunoblot with an AID antibody. Data shown are representative of three experiments.

switch transcripts, specifically abrogated CSR to IgG1, even though transcription through S γ 1 was unaffected (Lorenz et al., 1995). Additionally, increasing levels of S α transcripts by expression from a plasmid enhanced CSR to IgA in a cell line (Müller et al., 1998). Furthermore, while neither the specificity of the interaction nor the physiological significance of the binding was ascertained, AID was shown to bind various RNA in vitro, including tRNA and RNA from insect cells (Bransteitter et al., 2003; Dickerson et al., 2003) suggesting that RNA:AID interactions might be relevant for CSR. Finally, AID was also shown to preferentially mutate small RNA genes when expressed in yeast in a manner that suggests RNA might have a role in its recruitment (Taylor et al., 2014). Interestingly, RNA-guided processes have been shown to regulate DNA rearrangements in ciliates (Mochizuki et al., 2002; Nowacki et al., 2008) and to localize proteins to specific parts of the genome to modify chromatin (Tsai et al., 2010; Zhao et al., 2008). Switch RNAs are complementary to the S region DNA templates and would be ideal to provide specificity as a targeting factor to guide AID to its target DNA (S regions) during CSR. Thus, these observations led us to examine the possibility that switch transcripts can serve as molecular guides that target AID to S region DNA during CSR.

RESULTS

Switch RNA Binds AID

For switch transcripts to serve as targeting factors for AID, we reasoned that switch RNAs and AID may interact and therefore examined this notion in a series of RNA pull-down assays. Puri-

fied biotinylated in vitro transcribed (IVT) RNAs were allowed to fold into secondary/tertiary structures and examined for their ability to interact with AID present in extracts of CH12 cells stimulated for CSR. The mouse CH12 B lymphoma cell line switches at a high frequency from IgM to IgA with anti-CD40, interleukin 4 (IL-4), and transforming growth factor β (TGF- β) (henceforth referred to as CIT) stimulation and has been used as a model system to study CSR (Nowak et al., 2011; Pavri et al., 2010). As sense and anti-sense transcription have been reported through the S regions (Perlot et al., 2008), both sense (forward, F) and anti-sense (reverse, R) switch transcripts were analyzed in this assay. Sense switch μ (S μ F) and switch α (S α F) RNAs specifically bound AID, while their anti-sense counterparts (S μ R, S α R) did not (Figure 1A). Neither brain cytoplasmic RNA (Bc1), which has been described to form complexes with proteins (Zalfa et al., 2003), nor a fragment of RNA derived from one of the GAPDH introns associated with AID (Figure 1A). In contrast, interaction of these transcripts with Apobec3, a member of the cytidine deaminase family, was not detected (Figure 1A). An RNA dot-blot assay showed that all RNAs were recovered by streptavidin beads at comparable levels (Figure 1A). To assess the interaction between switch RNA and AID in vivo, we generated CH12 cells that stably express S α transcripts in either orientation (S α F or S α R) with an S1 tag, an RNA aptamer with affinity for streptavidin (Srisawat and Engelke, 2001). S1-tagged RNA from CIT-stimulated cells was recovered on streptavidin beads and probed for AID (Figure 1B). AID only co-purified with S1S α F RNA, even though more S1S α R RNA was recovered (Figure 1B). Apobec3 did not co-purify with either of the S α

transcripts (Figure 1B). Thus, switch RNA can bind to AID both in vitro and in vivo.

To determine the binding affinity between AID and switch RNAs, recombinant maltose-binding protein (MBP)-tagged mouse AID was purified and its binding to S μ F RNA was examined. Consistent with the interactions observed with cell extracts, in vitro binding assays showed that MBP-AID associated with S μ F (Figure 1C), indicating that the AID:RNA interaction is direct. As switch RNA molecules appeared to form large extensive structures that did not migrate into even low percentage gels, K_d values for the interaction between AID and RNAs could not be determined using conventional gel-shift assays. We therefore examined the relative binding affinities in competition assays. Inhibition of binding can be achieved in the nanomolar range (S μ F, ~4–5 nM; S α F, ~5–50 nM) (Figure 1C), suggesting that these interactions are highly specific and would likely have a physiological role.

G-Quadruplex Structures in Switch RNA

Interestingly, AID also co-purified with telomeric RNA (Figure 1A), which like switch transcripts, is G rich and consists of repetitive sequences (Azzalin and Lingner, 2007). Competitive binding experiments indicate that the affinity of AID for telomeric RNA is comparable to the binding affinities for sense switch RNAs (relative affinities, S μ F > telomere > S α F) (Figure 1C). This suggests that AID can associate with sense switch RNAs and similarly G-rich, repetitive RNAs through secondary and/or tertiary structures that are common to these transcripts. Nucleic acid sequences that are rich in guanine tracts can form G-quadruplex structures, wherein guanine residues are arranged in a planar conformation through Hoogsteen base-pairing interactions to form a four-stranded structure that is highly stabilized by a central K⁺ ion, but to a much lesser extent by Li⁺ (Lane et al., 2008; Sen and Gilbert, 1988; Williamson et al., 1989). Indeed, similar to telomeric RNA, computational analysis indicates that sense switch RNA sequences have strong G-quadruplex forming potential, while their anti-sense counterparts Bc1 and GAPDH intron do not (Figure 2A).

To assess whether switch RNAs can form these higher-order G-quadruplex structures, we used a synthetic RNA oligonucleotide representing four S μ repeats in tandem (S μ 4G) (Figure 2B). Analysis of S μ 4G mobility by gel electrophoresis showed that it migrated as a single species under denaturing conditions but as a higher molecular weight smear under native conditions (Figure 2C), indicating the formation of higher-order RNA structures. These higher-order structures were lost when S μ 4G was folded in the presence of Li⁺ (Figure 2C). Similar to what was observed for the longer ~1 kb S μ F IVT RNA (Figure 1A), S μ 4G interacted with AID from extracts of CIT-stimulated CH12 cells in the RNA-binding assays (Figure 2D). In comparison, no binding of S μ 4G to Apobec3 was observed in this assay (Figure 2D). Strikingly, when S μ 4G was folded in the presence of Li⁺, the interaction with AID was lost. In addition, a mutant form of S μ 4G containing G-to-C mutations that abolished the tandem guanine tracts (S μ 4Gmut) (Figure 2B) also did not bind AID (Figure 2D). Furthermore, in CH12 cells that stably expressed either S1-tagged S μ 4G or S μ 4Gmut transcripts, endogenous AID interacted with S μ 4G but not with S μ 4Gmut RNA (Figure 2E). Taken

together, these results suggest that switch RNAs can form G-quadruplexes and may bind AID through these structures.

To further demonstrate the ability of switch RNA to form G-quadruplex structures, we performed circular dichroism spectroscopy. S μ 4G, but not S μ 4Gmut, displayed the characteristic spectrum of parallel G-quadruplexes (Kumari et al., 2007; Xu et al., 2008), with a positive peak at ~260 nm and a negative peak at ~240 nm, both of which are reduced in the presence of Li⁺ (Figure 2F). These results are consistent with that observed for the well-characterized G-quadruplex-forming C9orf72 hexanucleotide repeat expansion (HRE) RNA (Haeusler et al., 2014) (Figure S1A). This signature spectrum was also evident in S μ F RNA indicating that it also forms parallel G-quadruplexes, while S μ R showed shifts in peaks to wavelengths of 270 and 233 nm that are not typical of these structures (Figure 2F). Formation of G-quadruplex structures was additionally verified in a ligand-binding colorimetric assay that is based on the ability of G-quadruplexes to bind hemin (Li et al., 2013). G-quadruplex:hemin complexes exhibit peroxidase-like activity that is detected as a green coloration when exposed to substrate (Figure S1B). This can be measured as an absorbance signal upon a spectral scan from 400 to 500 nm (Li et al., 2013) with a maxima at 420 nm as seen for the known G-quadruplex-forming C9orf72 HRE RNA (Haeusler et al., 2014) (Figure S1C). In this assay, S μ 4G RNA, when folded in the presence of K⁺ also showed the characteristic absorbance maxima at 420 nm, which was clearly reduced when S μ 4G RNA was folded in the presence of Li⁺ (Figure 2G). S μ 4Gmut RNA did not exhibit any detectable absorbance signal above the buffer control (Figure 2G). Similar to S μ 4G, S μ F also exhibited the characteristic absorbance maxima at 420 nm that was greatly reduced for S μ R (Figure 2G). Finally, the biotinylated S μ 4G used in the in vitro interaction experiments also exhibited G-quadruplex forming ability (Figures S1D and S1E). Taken together, these data strongly suggest that switch RNA can form G-quadruplexes and these structures might mediate the interaction with AID.

AID(G133V) Does Not Bind Switch RNA and Cannot Mediate CSR

Next, we investigated the G-quadruplex binding domain in AID. No RNA-binding domain in AID has been described; however, sequence alignment revealed that amino acids 130–138 of mouse AID shares some sequence homology with the RNA-binding domain of the well-characterized G-quadruplex RNA-binding protein RHAU (Creacy et al., 2008; Vaughn et al., 2005) (Figure S2A). Mutations of the two glycine residues to proline in RHAU greatly reduced its binding to G-quadruplex RNA (Lattmann et al., 2010). Likewise, mutation of the corresponding glycine-133 and glycine-137 in mouse AID to prolines completely abolished the ability of AID to restore CSR when expressed in AID-deficient mouse splenic B cells (Figures S2B–S2D). Furthermore, a G133V mutation in AID has been identified in two hyper-IgM patients with severe CSR defects (Mahdavi et al., 2012). As proline mutations can be disruptive to overall protein structure, and that the glycine-133 residue is conserved in AID across all species, mouse AID with the less bulky G133V mutation was characterized further.

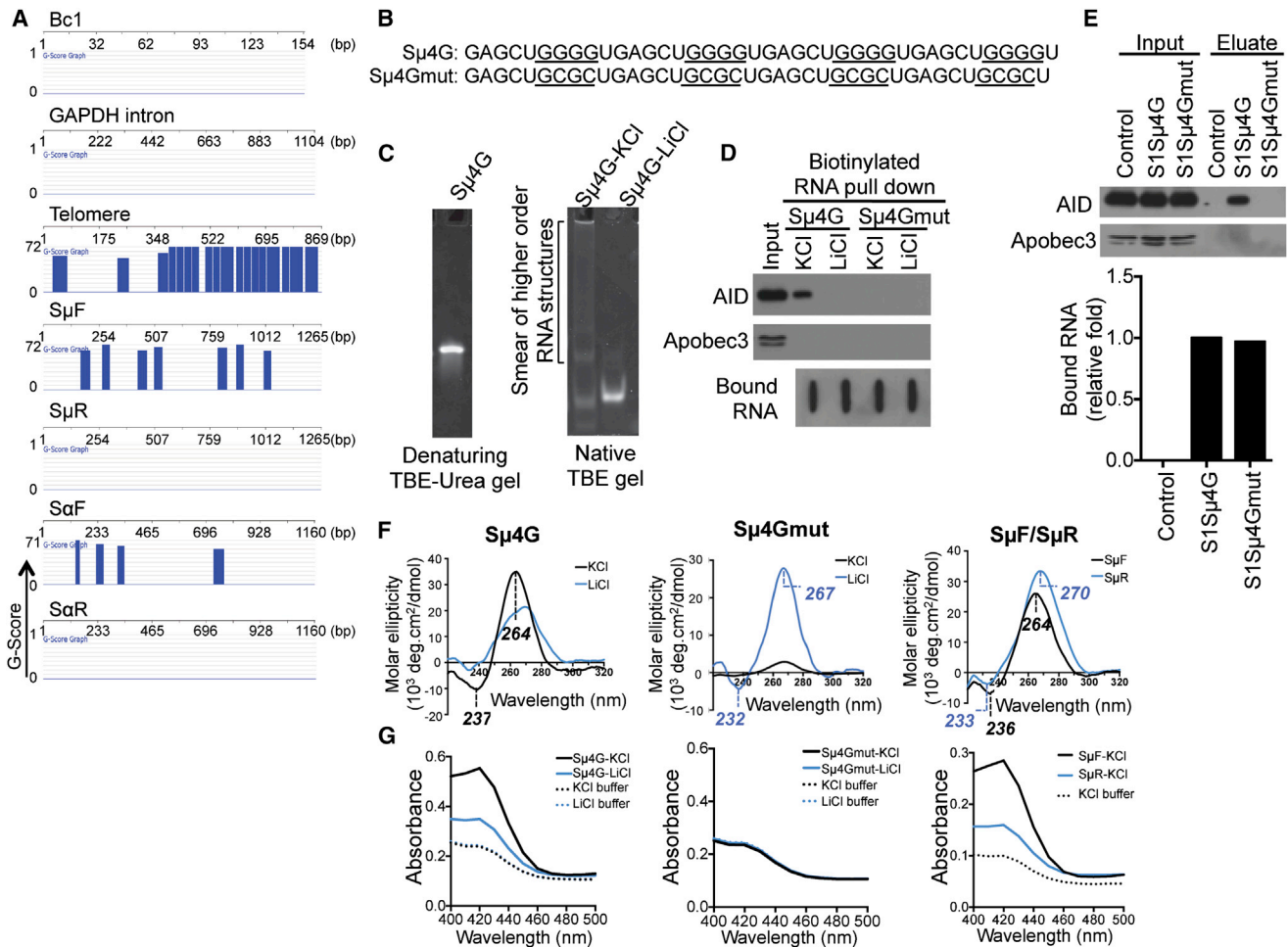


Figure 2. G-Quadruplex Structures in Switch RNA

(A) Sense switch RNAs and telomere RNA are predicted to form G-quadruplex structures. QGRS Mapper software was used to assess the G-quadruplex forming potential of the RNA sequences. The probability of G-quadruplex structure formation is reported as a G-score and represented over the corresponding regions in blue.

(B) Sequences of synthetic RNA oligonucleotides. Sμ4G, four Sμ repeats in tandem; mutant Sμ4G (Sμ4Gmut), with G-to-C mutations that abolish guanine tracts. Guanine tracts in Sμ4G and corresponding regions in Sμ4Gmut are underlined.

(C) Sμ4G was resolved on a denaturing gel, or folded in either KCl- or LiCl-containing buffer and resolved on a native gel, and stained with SYBR GOLD following electrophoresis.

(D) Sμ4G associates with AID while Sμ4Gmut does not. RNA pull-down was performed with biotinylated Sμ4G and Sμ4Gmut (folded in either KCl- or LiCl-containing buffer) and stimulated CH12 extracts. Recovered proteins were analyzed by immunoblot and bound RNAs by streptavidin-HRP blot. Result shown is representative of three independent pull-down experiments.

(E) Sμ4G interacts with AID in vivo. CH12 cells stably expressing S1-aptamer tagged Sμ4G or Sμ4Gmut transcripts were stimulated and ribonucleoprotein complexes were isolated on streptavidin beads. CH12 cells expressing anti-sense S1Sμ4G, which is unable to bind streptavidin, were used as control. RNA in the eluates was analyzed by RT-qPCR for amounts of exogenous transcripts relative to S1Sμ4G, while proteins in the eluates were analyzed by immunoblot. Result shown is representative of two independent pull-down experiments.

(F) Circular dichroism (CD) analysis of G-quadruplex structures. CD spectra of Sμ4G, Sμ4Gmut, SμF, and SμR RNAs (folded in either KCl- or LiCl-containing buffer). Wavelengths of observed peaks are indicated. Peaks characteristic of parallel G-quadruplexes are (positive, ~260 nm; negative, ~240 nm).

(G) Colorimetric assay of G-quadruplexes. RNAs were folded and incubated with hemin. G-quadruplexes bind hemin and the resultant complex exhibits peroxidase-like activity, which can be detected as an increase in absorbance around 420 nm when substrate is added (Haeusler et al., 2014; Li et al., 2013).

See also Figure S1.

Recombinant MBP-tagged wild-type (WT) and G133V mouse AID were purified (Figure 3A) and assessed for their ability to interact with switch RNAs. MBP-AID(G133V) was substantially impaired in its ability to bind switch RNAs compared to MBP-AID(WT) (Figure 3B), providing further support that the AID:RNA

interaction is specific and direct. To determine whether the failure to bind switch RNA is due to a general misfolding of MBP-AID(G133V), we carried out deamination assays on single-stranded DNA (ssDNA) substrates. While mouse AID is known to display weak deaminase activity in vitro (Chaudhuri

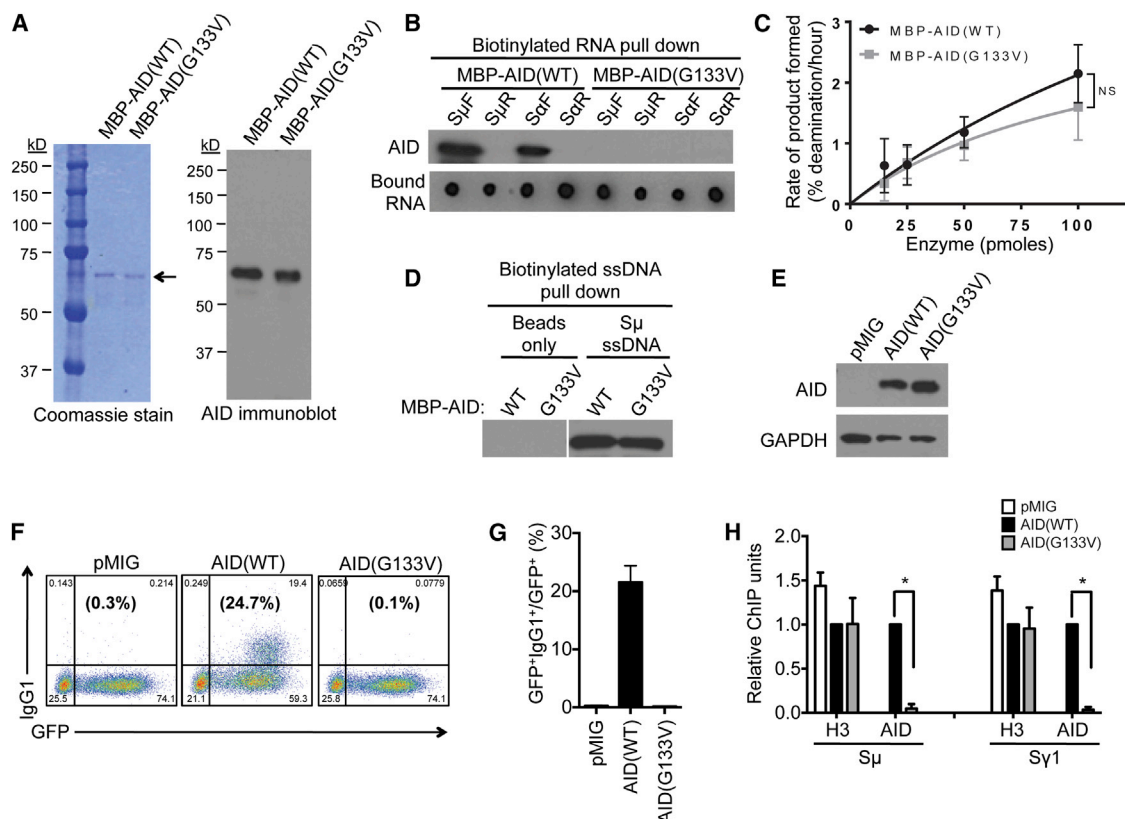


Figure 3. Glycine-133 of AID Is Critical for RNA Binding and CSR

(A) Purified recombinant MBP-tagged, wild-type (WT), or mutant AID (G133V) proteins were analyzed by Coomassie stain and immunoblot with AID antibody. Arrow indicates the size corresponding to full-length MBP-AID proteins on the Coomassie-stained gel.

(B) Purified proteins were used in the RNA pull-down assay with IVT biotinylated switch RNAs, followed by analysis of recovered proteins by immunoblot with an AID antibody, and bound RNAs by streptavidin-HRP dot blot. Result shown is representative of three independent pull-down experiments.

(C) Enzymatic activity of purified proteins was determined by a deamination assay. The rate of deamination was determined as a function of protein concentration as described in [Experimental Procedures](#). The average \pm SD of three independent protein preparations is shown; NS, $p =$ not significant, $p \geq 0.05$ at 25-, 50-, and 100-pmol enzyme concentrations.

(D) The G133V mutation does not affect binding of AID to single-stranded DNA (ssDNA). Purified proteins were incubated with biotinylated ssDNA, followed by pull-down with streptavidin beads and analysis by immunoblot with an AID antibody. Data shown are representative of three experiments.

(E–H) Splenic B cells were isolated from AID^{−/−} mice and transduced with vector control (pMIG) or vectors to express AID(WT) or AID(G133V).

(E) Expression of AID proteins was verified by immunoblot with AID or GAPDH (control) antibodies.

(F) CSR to IgG1 was determined by flow cytometry. A representative experiment is shown. The numbers in the corners indicate the percentage of cells in each quadrant, while the numbers in parentheses indicate the percentage of IgG1⁺ cells within the GFP⁺ gate.

(G) The average percentage of IgG1⁺ cells within the GFP⁺ gate from three independent experiments \pm SD is shown.

(H) Localization of AID proteins to S regions was determined by ChIP, using AID or H3 antibodies. S_μ and S_{γ1} DNA in ChIP samples was measured by qPCR, normalized to input DNA and the AID(WT) control. The mean of three independent experiments \pm SD is shown.

See also [Figure S2](#).

[et al., 2003](#)), the activity of MBP-AID(G133V) was comparable to MBP-AID(WT) ([Figure 3C](#)). Additionally, the ssDNA binding ability of MBP-AID(G133V) was similar to the wild-type protein ([Figure 3D](#)). Thus, the defect in RNA binding is unlikely to be due to general loss of the structural integrity of the MBP-AID(G133V) protein.

To determine the functional relevance of AID(G133V) in CSR, the mutant protein was expressed in AID-deficient splenic B cells. AID(G133V) was expressed at similar levels as AID(WT) ([Figure 3E](#)) and was present in the nucleus at comparable levels ([Figure S2E](#)), once again suggesting that the mutation did not have a gross effect on protein structure. Strikingly, AID(G133V)

was completely inactive in inducing CSR in AID-deficient B cells. While AID(WT) resulted in CSR in over 20% of the transduced cells, CSR in AID(G133V)-expressing cells was similar to that of the vector control ([Figures 3F and 3G](#)). Expression of AID(G133V) did not adversely affect the level of germline switch transcripts compared to AID(WT) ([Figure S2F](#)). To assess whether impaired CSR is due to reduced ability of AID(G133V) to bind DNA, chromatin immunoprecipitation (ChIP) was carried out with an AID antibody that can immunoprecipitate both wild-type and mutant proteins equally ([Figure S2G](#)). The ChIP experiments showed that AID(G133V) was significantly impaired in its ability to bind S region DNA ([Figure 3H](#)). While AID(WT) efficiently

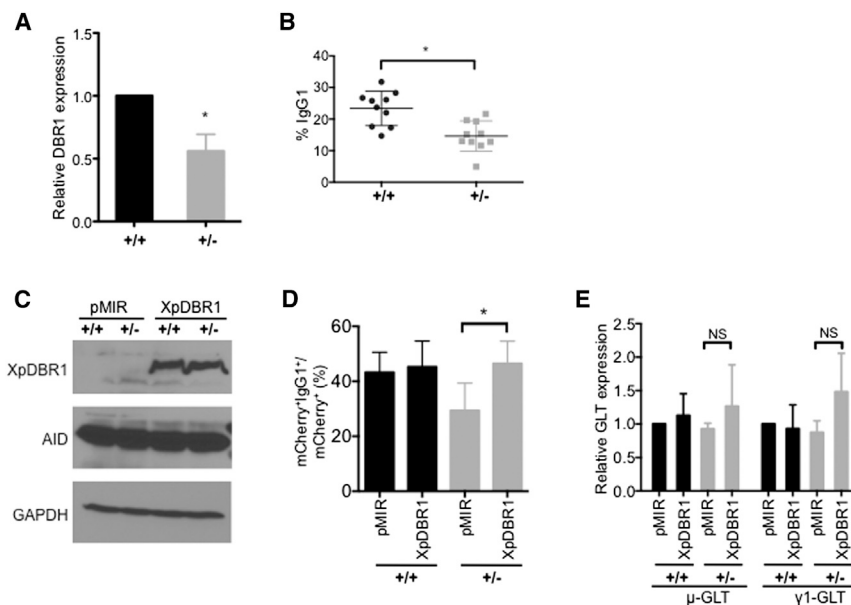


Figure 4. DBR1^{+/-} Splenic B Cells Exhibit Reduced CSR upon Ex Vivo Stimulation

Splenic B cells were isolated from heterozygous (^{+/-}) DBR1 gene-trapped mice and wild-type (^{+/+}) littermate controls, and stimulated in culture with anti-CD40 and IL-4 for 72 hr.

(A) Expression of full-length DBR1 is reduced in DBR1^{+/-} B cells. The level of DBR1 mRNA was measured by RT-qPCR using primers downstream of the gene trap insertion, normalized to β-actin mRNA and DBR1^{+/+} control. The average of four pairs of DBR1^{+/-} mice and their littermate DBR1^{+/+} controls ±SD is shown; *p < 0.05.

(B) CSR to IgG1 was determined at 72 hr post-stimulation by flow cytometry. Data show the mean of ten pairs of DBR1^{+/-} mice and their DBR1^{+/+} littermates ±SD. *p < 0.05.

(C–E) Splenic B cells from DBR1^{+/+} and DBR1^{+/-} mice were transduced with retroviral vector control (pMIR) or vector expressing Xpress-tagged DBR1 (XpDBR1).

(C) Expression of XpDBR1 and AID were verified by immunoblot by anti-Xpress, anti-AID, and anti-GAPDH (control) antibodies.

(D) CSR to IgG1 was determined by flow cytometry. The average percentage of IgG1⁺ cells within the mCherry⁺ gate from three independent experiments ±SD is shown. *p < 0.05.

(E) Expression of XpDBR1 does not adversely affect levels of μ- and γ1-germline switch transcripts (GLTs) compared to pMIR. Levels of μ- and γ1-GLT were determined by qRT-PCR, normalized to β-actin mRNA and pMIR control. Data represent the average of three independent experiments ±SD; NS, p = not significant, p ≥ 0.05.

See also Figure S3.

associated with S_μ and S_{γ1} (the two S regions that recombine upon anti-CD40 plus IL-4 stimulation), binding of AID(G133V) to these S regions was profoundly reduced (Figure 3H). Neither the wild-type, nor mutant AID protein associated with the control non-target locus C_{γ1} region (Figure S2H). The binding of histone H3 remained unaltered (Figure 3H; Figure S2H), demonstrating the specificity of the ChIP assays. This indicates that the failure of AID(G133V) to mediate CSR is due to a loss of binding to S region DNA. It is interesting to note that despite the abundance of telomeric RNA in splenic B cells (Figure S2J) and the ability of AID to interact with telomeric RNA in vitro (Figure 1A), AID did not bind telomere DNA (Figure S2I), probably because the telomeres are protected by a large number of proteins and tightly packed into heterochromatin, which might render them inaccessible to AID. Overall, the observation that a point mutation in AID that impairs its ability to bind G-quadruplex RNA also markedly reduces its ability to bind DNA strongly supports the notion that switch RNAs can guide AID to DNA.

Debranching of RNA Lariats Is Required for CSR

The direct association of AID with switch (or “guide”) RNAs, and the defect in CSR when the interaction is impaired, led us to hypothesize that perturbations to the processing of germline switch transcripts could interfere with the generation of the guide RNAs that recruit AID to S region DNA. As S region sequences lie within the intronic region of the germline switch transcripts, we tested this hypothesis by depleting the lariat debranching enzyme, DBR1, to perturb the processing of the switch RNAs without affecting transcription and splicing, upstream events that are known to be important for CSR (Matthews et al., 2014). DBR1 is

responsible for debranching intronic lariats by catalyzing the hydrolysis of the 2',5'-phosphodiester bond at the branchpoint (Arenas and Hurwitz, 1987; Ruskin and Green, 1985). We generated DBR1 “gene-trapped” mice in which a truncated, non-functional, fusion protein, which is missing ~80% of the DBR1 protein, is produced (Figure S3A). No live homozygous (DBR1^{-/-}) mutant mice were obtained from breeding DBR1^{+/-} mice, indicating that DBR1 is required for embryonic development. However, DBR1^{+/-} splenic B cells activated for CSR exhibited a significant decrease in expression of full-length DBR1 mRNA (Figure 4A) with concomitant expression of the gene-trapped fusion mRNA (Figure S3B). DBR1^{+/-} B cells expressed similar levels of AID protein (Figure S3C) and germline transcripts (Figure S3D), and proliferated at comparable rates (Figures S3E and S3F) to DBR1^{+/+} littermate controls. DBR1^{+/-} B cells stimulated ex vivo showed a small but significant reduction in CSR to IgG1 (Figure 4B; Figure S3G). The CSR defect was rescued by expression of Xpress-tagged DBR1 (XpDBR1) (Figures 4C and 4D). Exogenous expression of XpDBR1 did not affect levels of AID (Figure 4C) or germline switch transcripts (Figure 4E).

Impaired CSR in DBR1^{+/-} B cells was accompanied by a slightly reduced, though not statistically significant, frequency of mutations in S_μ (Figure S3H). Somatic hypermutation (SHM) in Peyer's patches B cells remained unaffected (Figure S3I). While these data suggest that DBR1 influences CSR, the relative difficulty of performing experimental manipulations in short-lived ex vivo cultured mouse splenic B cells, as well as the modest CSR and S_μ mutation frequency defects in DBR1^{+/-} mice, led us to use the CH12 cell line as a model system to further elucidate the roles of switch RNAs in CSR.

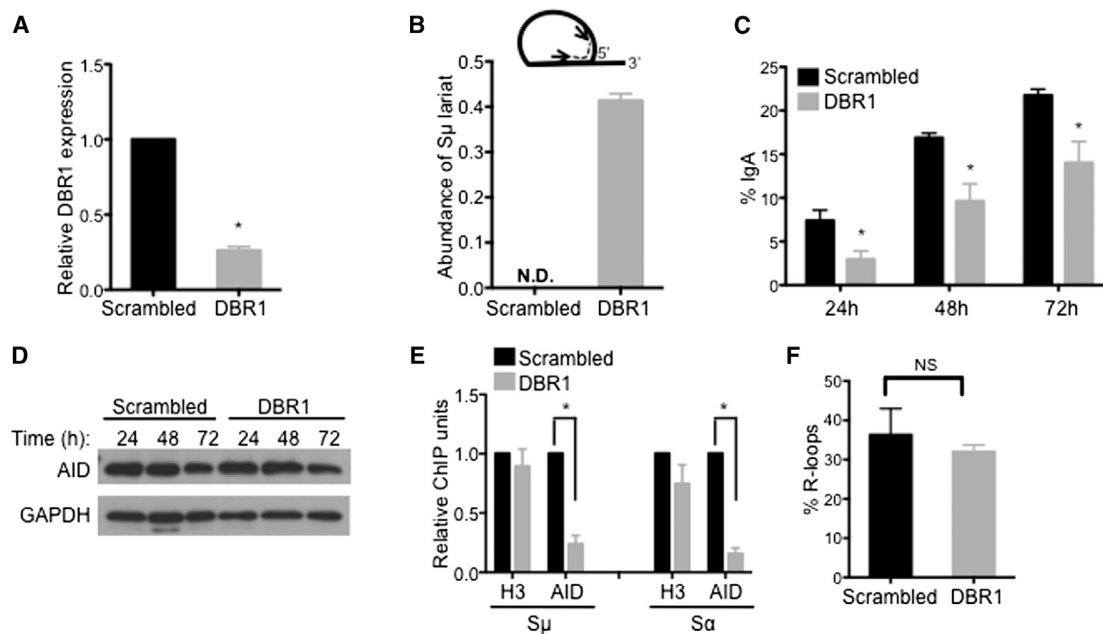


Figure 5. Debranching of Intronic RNA Lariats Is Required for Targeting of AID to S Regions and Efficient CSR

(A–F) Knockdown of DBR1 impairs CSR and AID localization to S regions. CH12 cells were transduced with shRNA against DBR1 or scrambled control shRNA. (A) Knockdown of DBR1 transcripts was determined by RT-qPCR.

(B) Accumulation of S μ lariats following DBR1 knockdown was determined by RT-qPCR using primers (arrows) positioned across the branchpoint as shown (inset). N.D., not detected.

Data in (A) and (B) were normalized to β -actin mRNA and the scrambled control; the average of at least three independent knockdown experiments \pm SD are shown; * $p < 0.05$.

(C) CSR to IgA was assayed by flow cytometry at indicated times following stimulation. Data show the mean of three independent knockdown experiments \pm SD. * $p < 0.05$.

(D) AID protein expression was analyzed at indicated times following stimulation by immunoblot with AID or GAPDH (control) antibodies.

(E) Localization of AID to S regions was determined by ChIP on cells 48 hr post-stimulation using AID or H3 antibodies. S μ and S α DNA in ChIP samples were measured by qPCR and normalized to input DNA and the scrambled control. Data represent mean of three independent knockdown experiments \pm SD. * $p < 0.05$.

(F) R-loop formation is unaffected by DBR1 knockdown. Genomic DNA was prepared from cells 24 hr post-stimulation and treated with the bisulfite modification assay. S α was examined and the number of molecules containing R-loops is represented as a percentage of the total number of DNA amplicons sequenced. The mean of three independent knockdown experiments \pm SD is shown; NS, $p =$ not significant, $p \geq 0.05$.

See also Figures S4 and S5.

Knockdown of DBR1 was achieved in CH12 cells using small hairpin RNA (shRNA) (Figure 5A) and was accompanied by an increase in the RNA lariat levels of S μ (Figure 5B; Figure S4B) and β -actin intron 3 (Figures S4A and S4C) as compared to the scrambled shRNA control, confirming a reduction in DBR1 enzymatic activity. When DBR1 knockdown cells were stimulated with CIT, we observed a significant reduction in CSR to IgA compared to control cells at all time points assayed by flow cytometry (Figure 5C) and by semiquantitative measurement of the I α -C μ circle transcripts produced from the excised extrachromosomal DNA (Figure S4D). Levels of mature germline transcripts were not altered in DBR1 knockdown cells (Figure S4E), suggesting that transcription and splicing were not affected. DBR1 knockdown did not affect expression of AID mRNA (Figure S4F) or protein (Figure 5D). DBR1 knockdown cells proliferated at rates comparable to the scrambled control (Figures S4G and S4H). Thus, DBR1 knockdown significantly impaired CSR without affecting transcription, splicing of primary germline switch transcripts, AID expression, or proliferation.

To determine whether the CSR defect in DBR1 knockdown CH12 cells is due to impaired targeting of AID to S region DNA, ChIP assays were performed. DBR1 knockdown led to a significant reduction in AID localization (\sim 5-fold) to both S μ and S α , while the binding of histone H3 was unaffected (Figure 5E). AID did not associate with the control non-target locus C γ 1 (Figure S5A), demonstrating the specificity of the assay. The defect in the ability of AID to bind S regions in DBR1 knockdown cells was not due to altered stability or abundance of R-loops at S region DNA (Kao et al., 2013) (Figure 5F; Figure S5B). The localization of RNA polymerase II and Spt5 to S regions were unaffected by DBR1 knockdown (Figure S5C), indicating that the loss of AID binding at S regions was not due to perturbations to these known factors of AID targeting (Pavri et al., 2010). These data suggest that post-transcriptional factors (guide RNAs) facilitate AID targeting to S regions in addition to previously characterized co-transcriptional factors. It is noteworthy that DBR1 knockdown occasionally resulted in a more severe defect in CSR (Figure S5D), which correlated with a larger reduction of AID localization (>100 -fold) to S regions (Figures S5E and S5F). Despite

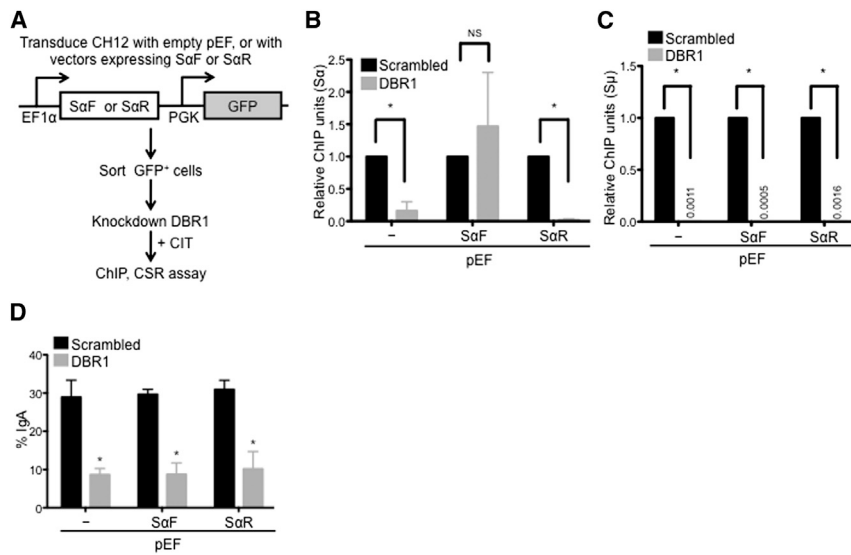


Figure 6. Localization of AID to S Regions Can Be Restored by Exogenous Expression of Switch Transcripts

(A) Experimental design. CH12 cells were transduced with empty vector (pEF) or vectors to express forward/sense or reverse/anti-sense switch α RNA (pEF-S α F or pEF-S α R, respectively). Transduced cells were sorted, infected with scrambled control shRNA or shRNA to knockdown DBR1, and stimulated with CIT.

(B and C) Expression of exogenous S α F rescues AID localization to S α at the endogenous IgH locus, but not to the non-complementary S μ . ChIP was performed on cells 48 hr post stimulation by immunoprecipitation with anti-AID antibodies. S α (B) and S μ (C) DNA in ChIP samples was quantified by qPCR and normalized to input and scrambled control.

(D) Expression of exogenous S α F does not rescue CSR in DBR1 knockdown cells. CSR to IgA was assayed 72 hr after stimulation by flow cytometry. Data in (B)–(D) represent the mean of three independent experiments \pm SD. * $p < 0.05$; NS, $p =$ not significant, $p \geq 0.05$. See also Figure S6.

the drastic loss of AID targeting in this instance, AID binding to S regions was not completely abrogated in DBR1 knockdown cells, as evident from the qPCR results that showed enrichment of S μ and S α fragments in anti-AID ChIPs as compared to the non-specific IgG control (Figure S5G). Although it is unclear why the magnitude of the defect varies, DBR1 knockdown consistently led to a reproducible reduction in AID localization to S region DNA.

Expression of Switch RNA in *trans* Rescues CSR Defect in DBR1-Depleted Cells

Given that DBR1 is responsible for debranching all intronic lariats in the cell, it remains to be determined whether the loss of AID targeting can be attributed to impaired processing of switch transcripts alone. To determine whether switch transcripts can guide AID to S region DNA during CSR, we expressed S α transcripts in either the sense (exoS α F) or anti-sense (exoS α R) orientation in CH12 cells and examined whether these exogenous switch transcripts could rescue AID localization to S regions and CSR in DBR1 knockdown cells (Figure 6A). Exogenous S α transcripts were readily detected (Figure S6A) and the cells expressed similar levels of AID upon CIT stimulation (Figure S6B). ChIP analyses showed that expression of exoS α F, but not exoS α R, restored binding of AID to endogenous S α DNA (Figure 6B), despite exoS α R being expressed to higher levels than exoS α F (Figure S6C). The qPCR quantification in the ChIP analyses was performed using primers that detected the endogenous S α locus, but not the exogenously transduced sequence, indicating that the observed results are a rescue of AID targeting to the endogenous S α locus. Interestingly, neither exoS α F nor exoS α R could rescue AID localization to the non-complementary endogenous S μ DNA in DBR1 knockdown cells (Figure 6C), suggesting that switch RNAs can target AID to S region DNA in a sequence-specific manner. As expected from the selective rescue of AID localization to only the S α DNA, exogenous expression of S α transcripts could not restore CSR to IgA (Figure 6D). Additionally, higher expression of exoS α R compared

to exoS α F (Figure S6C) was not detrimental to CSR, as cells expressing exoS α R switched at a level comparable to cells expressing exoS α F and those infected with the empty vector control (Figure 6D).

The inability of exogenous S α alone to rescue IgA levels suggests that restoration of AID targeting to both participating switch loci, S μ and S α , is required before productive CSR to IgA can occur. To test this hypothesis, we transduced CH12 cells to co-express sense (forward: exoS μ F + exoS α F) or anti-sense (reverse: exoS μ R + exoS α R) transcripts of S μ and S α (Figure 7A). Cells expressing exogenous switch transcripts were then knocked down for DBR1 expression (Figure S6D). DBR1 depletion increased lariat accumulation (Figure S6E) but did not affect steady-state levels of exogenous switch transcripts (Figure S6F) or AID expression (Figure S6G). Strikingly, co-expression of exoS μ F and exoS α F transcripts significantly rescued switching to IgA in DBR1 knockdown cells (Figure 7B). In contrast, dual expression of exoS μ R and exoS α R transcripts had no effect in restoring IgA levels in DBR1 knockdown cells (Figure 7B). Thus, expression of sense S μ and S α transcripts together can functionally reconstitute CSR in cells where debranching of intronic RNA lariats was inhibited. It is interesting to note that co-expression of S μ F and S α F transcripts was unable to increase switching to IgA in the absence of DBR1 knockdown (Figure 7B). This suggests that endogenous guide RNAs are not limiting during CSR, or perhaps the rate of the switching reaction is at a maxima and the supply of exogenous guide RNAs could not increase CSR to IgA any further in these cells. Taken together, these data provide strong experimental support for a model (Figure 7C) wherein switch RNA, through its ability to fold into G-quadruplex structures, can bind AID and target AID to S region DNA during CSR.

DISCUSSION

Collectively, our results provide strong evidence for the existence of an RNA cofactor in the targeting of AID to S regions

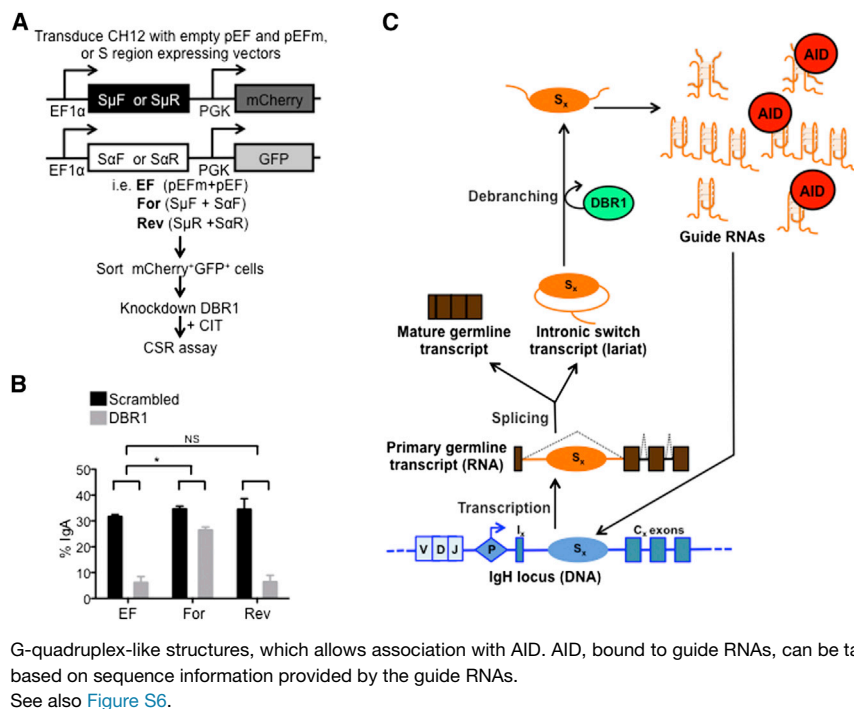


Figure 7. Co-expression of Sense S_{μ} and S_{α} Transcripts Rescues CSR to IgA in DBR1 Knockdown Cells

(A) Experimental design. CH12 cells were transduced with empty vectors (EF) or vectors expressing forward/sense (For) or reverse/anti-sense (Rev) switch transcripts. Doubly transduced cells were sorted, infected with scrambled control shRNA or shRNA to knockdown DBR1, and stimulated with CIT.

(B) CSR to IgA was assayed by flow cytometry 72 hr after stimulation. The average of three independent experiments \pm SD is shown. * $p < 0.05$; NS, $p =$ not significant, $p \geq 0.05$.

(C) Model for RNA-mediated targeting of AID during CSR. When B cells are stimulated to undergo class switching, transcription occurs at each of the recombining S regions to produce primary switch transcripts. Primary transcripts are spliced, with the intronic switch region sequences (S_x) spliced out as a lariat intermediate. Debranching enzyme (DBR1) catalyzes the release of the lariat and debranches the switch transcript into its linear form. The linear switch transcript, free of exonic sequences, can function as guide RNAs by forming G-quadruplex or

G-quadruplex-like structures, which allows association with AID. AID, bound to guide RNAs, can be targeted specifically to the complementary S region DNA based on sequence information provided by the guide RNAs. See also Figure S6.

during CSR. In this model, guide RNAs derived from the intronic region of germline switch transcripts can form G-quadruplexes or G-quadruplex-like RNA structures that allow association with AID, thereby guiding AID preferentially to the complementary S region DNA in a sequence-specific manner (Figure 7C). Identification of two patients with hyper-IgM syndrome, who carry the G133V mutation (Mahdavian et al., 2012) in the putative G-quadruplex RNA-binding domain of AID that disrupts binding to switch RNAs, further highlights the relevance of the switch RNA:AID interaction in CSR.

The importance of sequence information encoded by guide RNAs for their function indicates that base-pairing-mediated recognition is likely involved. Yet, according to the prevailing R-loop-based model for CSR (Chaudhuri et al., 2007), the template DNA strand of S regions is stably hybridized to the nascent primary transcript. Thus, the interaction of the RNA:AID complex with the template DNA strand likely requires displacement of the nascent transcripts by the guide RNA molecules. Transcription through the switch regions may be temporally regulated, with a wave of transcription to generate the guide RNAs, followed by a period of transcriptional quiescence at the locus to free the template strand to interact with guide RNAs bound to AID molecules. Alternatively, anti-sense transcription (Perlot et al., 2008) through the IgH locus could expose the sense strand to base pair with guide RNAs. Finally, it is also possible that following R-loop collapse (maybe after RNaseH action or RNA exosome activity (Basu et al., 2011)), the cDNA strands misalign due to the repetitive nature of S regions and results in exposed stretches of ssDNA (Yu and Lieber, 2003), providing an ideal seed sequence for guide RNA binding. RNA:RNA base-pairing could also allow guide RNAs to find the S region DNA. Interestingly, the anti-sense switch tran-

scripts appear to be inert in that, while they do not promote CSR, they do not act as decoys and decrease CSR (Figure 7B). This suggests that when expressed in *trans*, anti-sense switch transcripts do not interact with or affect the activity of the guide RNAs and sense germline transcripts. Nevertheless, the possibility remains that anti-sense transcripts may have a role in *cis*, and that nascent anti-sense transcripts at the switch locus could serve as docking sites for guide RNAs to find their target DNA region. Alternatively, guide RNAs may interact with the nascent sense switch transcripts, perhaps by the same interactions that allow switch RNAs to form G-quadruplex structures (Figures 2F and 2G), and thus be localized to the vicinity of their target DNA.

RNA-mediated recruitment of AID to regions outside the immunoglobulin locus may also have implications for the aberrant targeting of AID to regions outside the immunoglobulin locus. As AID has the ability to associate with RNA by binding G-quadruplex structures (Figures 1 and 2), AID could potentially associate with other cellular RNA that fold into similar structures, thus mistargeting AID to these other genomic loci. For instance, c-Myc has been reported to be a hotspot for AID activity outside the immunoglobulin locus and contributes to the c-Myc/IgH translocations seen in B cell lymphomas (Nussenzweig and Nussenzweig, 2010). G-quadruplex structures in DNA have been implicated in CSR (Dempsey et al., 1999) and have also been found in the first exon and intron of c-Myc, which correspond to common breakpoints in c-Myc/IgH translocations that involve AID (Duquette et al., 2004, 2005). Interestingly, examination of a subset of genes that are bound by AID in ChIP sequencing (ChIP-seq) analysis (Yamane et al., 2011) showed that the primary transcripts derived from these genes have greater potential to form G-quadruplexes than RNA transcribed from genes that are not targeted by AID

(Figure S7). While AID activity at these non-Ig loci has been attributed to secondary structures in the DNA, it is tempting to speculate that the G-quadruplex RNA molecules derived from this region mediates mistargeting of AID. Additionally, recent studies further revealed that anti-sense and convergent transcription at super-enhancers, especially at these non-Ig hot-spots, can facilitate mistargeting by providing single-stranded DNA substrates for AID (Meng et al., 2014; Pefanis et al., 2014; Qian et al., 2014). The high binding affinity of AID for switch RNA that is in the nanomolar range could potentially facilitate AID to distinguish the switch RNAs from the other RNA species in the physiological setting of the cell. A detailed landscape of transcriptome-wide association of AID will be important to better establish a global map of AID:RNA interactions. However, these studies await generation of crosslinking and immunoprecipitation of RNA-protein complex sequencing (CLIP-seq) or RNA immunoprecipitation sequencing (RIP-seq) grade antibodies and better sequencing approaches to map repetitive sequences before such experiments can be meaningfully undertaken.

Recruitment of AID to S regions by interaction with RNA polymerase II in a co-transcriptional step and by binding to switch RNA molecules in a post-transcriptional step provides two distinct mechanisms by which AID could be efficiently and specifically targeted to the *IgH* locus during CSR. This two-step recruitment not only ensures the localization of a high density of AID molecules at S regions required for CSR but also provides a means by which this general mutator is sequestered from other regions of the genome. In this regard, S regions in *Xenopus* are not G rich and RNA transcribed from the *Xenopus IgH* locus is not predicted to form G-quadruplex structures. It is likely that CSR in *Xenopus* occurs through a SHM-like mechanism that does not require defined RNA structures (Zarrin et al., 2004).

In summary, we have uncovered a novel mechanism for the targeting of AID specifically to S regions at the *IgH* locus during CSR that is based on sequence information encoded in RNA. This study specifies a role for germline transcripts independent of transcription and provides an explanation for the long-standing link between the requirement of splicing and CSR (Lorenz et al., 1995). RNA-guided processes are emerging as an efficient mechanism to target proteins to defined genomic regions (Tsai et al., 2010; Zhao et al., 2008), and our findings reveal CSR to be yet another example of this versatile system.

EXPERIMENTAL PROCEDURES

Mice

AID^{-/-} mice were a kind gift from Dr. T. Honjo. DBR1^{+/-} mice were generated at MSKCC transgenic mouse core facility from gene-trapped embryonic stem cell line YTA280 (strain 129P2/OlaHsd) obtained from BayGenomics/Mutant Mouse Regional Resource Centers. All animals were housed according to the guidelines for animal care of MSKCC Research Animal Resource Center.

RNA Folding and In Vitro RNA Pull-Down Assay

RNA folding and pull-down assay were performed as described in Booy et al. (2012). Briefly, RNAs were folded by heating at 95°C and then allowed to cool passively to room temperature. Purified AID proteins or whole-cell extracts from stimulated CH12 cells were incubated with folded biotinylated RNAs,

followed by pull-down with streptavidin beads. Bound proteins and RNAs were analyzed by immunoblot and dot blot, respectively. For details, see Extended Experimental Procedures.

Purification of S1-Tagged Ribonucleoprotein Complexes

Whole-cell extracts were prepared from stimulated CH12 cells that stably express S1-tagged switch RNAs. S1-tagged RNAs and associated proteins were recovered by pull-down with streptavidin beads and analyzed by RT-qPCR and immunoblot, respectively. See Extended Experimental Procedures for more details.

Deamination Assay

Deamination assay was performed as described (Nabel et al., 2012), using a 5'-radiolabeled 30-bp oligonucleotide with a single cytosine as substrate. Following incubation with MBP-AID proteins, UDG was added, and DNA at abasic sites was cleaved by heating in 0.1 N NaOH. Samples were resolved and analyzed by autoradiography; percentage product formed over time was calculated and normalized to the 0-hr control. The rate of the reaction was calculated from the slope of the curve and plotted against concentration of enzyme. See Extended Experimental Procedures for more details.

Computational Analysis of RNA Sequences

The G-quadruplex prediction software QGPR Mapper (<http://bioinformatics.ramapo.edu/QGRS/analyze.php>) (Kikin et al., 2006) was used to assess the G-quadruplex forming potential of RNA sequences. Parameter used are as follows: max length = 45, minimum G-group size = 3, loop size = 0–36. The highest G-score for each primary transcript was noted as the G-score for that sequence.

Circular Dichroism Spectroscopy

RNA oligonucleotides (C9orf72 HRE, S_μ4G and S_μ4Gmut) were folded at 10 μM, while S_μF and S_μR RNA were folded at 0.5 μM. Circular dichroism (CD) spectra were obtained using an Aviv 202 CD spectrometer 62DS at 25°C, with wavelength scan range of 220–320 nm and path length of 1 mm. Spectra were subtracted for buffer controls, and smoothing was performed using Prism software by averaging four neighboring points.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and two tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.03.020>.

AUTHOR CONTRIBUTIONS

S.Z., B.Q.V., and J.C. developed ideas, designed experiments, and wrote the manuscript. S.Z. performed most of the experiments described in the manuscript. B.Q.V. carried out some of the ChIP assays, retroviral reconstitution, SHM, and protein fractionation studies. B.V. purified MBP-AID proteins and performed the XpDBR1 experiments, deamination, and colorimetric assays. J.-Y.L. and F.-T.H. carried out the R-loop assays.

ACKNOWLEDGMENTS

The authors wish to thank T. Honjo for AID^{-/-} mice and CH12 cells, S. Kabir (Rockefeller University) for telomere-specific probes, and S. Vasudevan (Harvard University) and T. Agarwal for their helpful technical advice. We thank members of J.C.'s laboratory for their comments and discussion. This work was supported by grants from the NIH (1R01AI072194) and the Starr Cancer Foundation (I4-A447 and I7-A767) to J.C.

Received: October 1, 2014

Revised: December 31, 2014

Accepted: February 27, 2015

Published: May 7, 2015

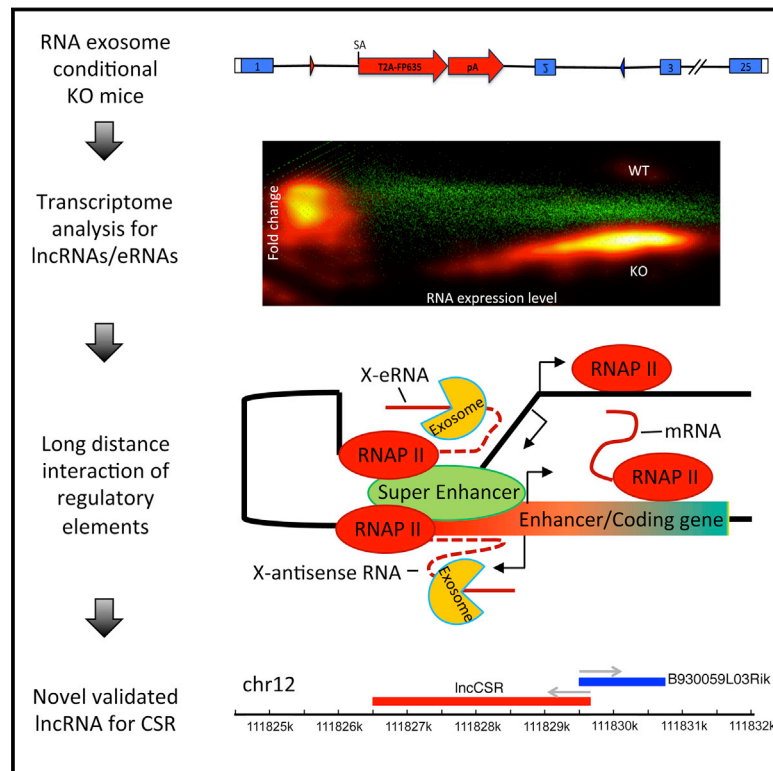
REFERENCES

- Arenas, J., and Hurwitz, J. (1987). Purification of a RNA debranching activity from HeLa cells. *J. Biol. Chem.* 262, 4274–4279.
- Azzalin, C.M., and Lingner, J. (2007). Molecular biology: damage control. *Nature* 448, 1001–1002.
- Basu, U., Meng, F.L., Keim, C., Grinstein, V., Pefanis, E., Eccleston, J., Zhang, T., Myers, D., Wasserman, C.R., Wesemann, D.R., et al. (2011). The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. *Cell* 144, 353–363.
- Booy, E.P., Meier, M., Okun, N., Novakowski, S.K., Xiong, S., Stetefeld, J., and McKenna, S.A. (2012). The RNA helicase RHAU (DHX36) unwinds a G4-quadruplex in human telomerase RNA and promotes the formation of the P1 helix template boundary. *Nucleic Acids Res.* 40, 4110–4124.
- Bransteitter, R., Pham, P., Scharff, M.D., and Goodman, M.F. (2003). Activation-induced cytidine deaminase deaminates deoxycytidine on single-stranded DNA but requires the action of RNase. *Proc. Natl. Acad. Sci. USA* 100, 4102–4107.
- Chaudhuri, J., Tian, M., Khuong, C., Chua, K., Pinaud, E., and Alt, F.W. (2003). Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature* 422, 726–730.
- Chaudhuri, J., Basu, U., Zarrin, A., Yan, C., Franco, S., Perlot, T., Vuong, B., Wang, J., Phan, R.T., Datta, A., et al. (2007). Evolution of the immunoglobulin heavy chain class switch recombination mechanism. *Adv. Immunol.* 94, 157–214.
- Creacy, S.D., Routh, E.D., Iwamoto, F., Nagamine, Y., Akman, S.A., and Vaughn, J.P. (2008). G4 resolvase 1 binds both DNA and RNA tetramolecular quadruplex with high affinity and is the major source of tetramolecular quadruplex G4-DNA and G4-RNA resolving activity in HeLa cell lysates. *J. Biol. Chem.* 283, 34626–34634.
- Dempsey, L.A., Sun, H., Hanakahi, L.A., and Maizels, N. (1999). G4 DNA binding by LR1 and its subunits, nucleolin and hnRNP D, A role for G-G pairing in immunoglobulin switch recombination. *J. Biol. Chem.* 274, 1066–1071.
- Dickerson, S.K., Market, E., Besmer, E., and Papavasiliou, F.N. (2003). AID mediates hypermutation by deaminating single stranded DNA. *J. Exp. Med.* 197, 1291–1296.
- Duquette, M.L., Handa, P., Vincent, J.A., Taylor, A.F., and Maizels, N. (2004). Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev.* 18, 1618–1629.
- Duquette, M.L., Pham, P., Goodman, M.F., and Maizels, N. (2005). AID binds to transcription-induced structures in c-MYC that map to regions associated with translocation and hypermutation. *Oncogene* 24, 5791–5798.
- Haeusler, A.R., Donnelly, C.J., Periz, G., Simko, E.A., Shaw, P.G., Kim, M.S., Maragakis, N.J., Troncoso, J.C., Pandey, A., Sattler, R., et al. (2014). C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* 507, 195–200.
- Kao, Y.P., Hsieh, W.C., Hung, S.T., Huang, C.W., Lieber, M.R., and Huang, F.T. (2013). Detection and characterization of R-loops at the murine immunoglobulin S α region. *Mol. Immunol.* 54, 208–216.
- Kikin, O., D'Antonio, L., and Bagga, P.S. (2006). QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.* 34, W676–W682.
- Kumari, S., Bugaut, A., Huppert, J.L., and Balasubramanian, S. (2007). An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nat. Chem. Biol.* 3, 218–221.
- Lane, A.N., Chaires, J.B., Gray, R.D., and Trent, J.O. (2008). Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.* 36, 5482–5515.
- Lattmann, S., Giri, B., Vaughn, J.P., Akman, S.A., and Nagamine, Y. (2010). Role of the amino terminal RHAU-specific motif in the recognition and resolution of guanine quadruplex-RNA by the DEAH-box RNA helicase RHAU. *Nucleic Acids Res.* 38, 6219–6233.
- Li, H., Liu, J., Fang, Y., Qin, Y., Xu, S., Liu, Y., and Wang, E. (2013). G-quadruplex-based ultrasensitive and selective detection of histidine and cysteine. *Biosens. Bioelectron.* 41, 563–568.
- Lorenz, M., Jung, S., and Radbruch, A. (1995). Switch transcripts in immunoglobulin class switching. *Science* 267, 1825–1828.
- Mahdavian, S.A., Hirbod-Mobarakeh, A., Wang, N., Aghamohammadi, A., Hammarström, L., Masjedi, M.R., Pan-Hammarström, Q., and Rezaei, N. (2012). Novel mutation of the activation-induced cytidine deaminase gene in a Tajik family: special review on hyper-immunoglobulin M syndrome. *Expert Rev. Clin. Immunol.* 8, 539–546.
- Matthews, A.J., Zheng, S., DiMenna, L.J., and Chaudhuri, J. (2014). Regulation of immunoglobulin class-switch recombination: choreography of noncoding transcription, targeted DNA deamination, and long-range DNA repair. *Adv. Immunol.* 122, 1–57.
- Maul, R.W., Saribasak, H., Martomo, S.A., McClure, R.L., Yang, W., Vaisman, A., Gramlich, H.S., Schatz, D.G., Woodgate, R., Wilson, D.M., 3rd, and Gearhart, P.J. (2011). Uracil residues dependent on the deaminase AID in immunoglobulin gene variable and switch regions. *Nat. Immunol.* 12, 70–76.
- Meng, F.L., Du, Z., Federation, A., Hu, J., Wang, Q., Kieffer-Kwon, K.R., Meyers, R.M., Amor, C., Wasserman, C.R., Neuberg, D., et al. (2014). Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. *Cell* 159, 1538–1548.
- Mochizuki, K., Fine, N.A., Fujisawa, T., and Gorovsky, M.A. (2002). Analysis of a piwi-related gene implicates small RNAs in genome rearrangement in tetrahymena. *Cell* 110, 689–699.
- Müller, J.R., Giese, T., Henry, D.L., Mushinski, J.F., and Marcu, K.B. (1998). Generation of switch hybrid DNA between Ig heavy chain-mu and downstream switch regions in B lymphocytes. *J. Immunol.* 161, 1354–1362.
- Muramatsu, M., Kinoshita, K., Fagarasan, S., Yamada, S., Shinkai, Y., and Honjo, T. (2000). Class switch recombination and hypermutation require activation-induced cytidine deaminase (AID), a potential RNA editing enzyme. *Cell* 102, 553–563.
- Nabel, C.S., Jia, H., Ye, Y., Shen, L., Goldschmidt, H.L., Stivers, J.T., Zhang, Y., and Kohli, R.M. (2012). AID/APOBEC deaminases disfavor modified cytosines implicated in DNA demethylation. *Nat. Chem. Biol.* 8, 751–758.
- Nambu, Y., Sugai, M., Gonda, H., Lee, C.G., Katakai, T., Agata, Y., Yokota, Y., and Shimizu, A. (2003). Transcription-coupled events associating with immunoglobulin switch region chromatin. *Science* 302, 2137–2140.
- Nowacki, M., Vijayan, V., Zhou, Y., Schotanus, K., Doak, T.G., and Landweber, L.F. (2008). RNA-mediated epigenetic programming of a genome-rearrangement pathway. *Nature* 451, 153–158.
- Nowak, U., Matthews, A.J., Zheng, S., and Chaudhuri, J. (2011). The splicing regulator PTBP2 interacts with the cytidine deaminase AID and promotes binding of AID to switch-region DNA. *Nat. Immunol.* 12, 160–166.
- Nussenzweig, A., and Nussenzweig, M.C. (2010). Origin of chromosomal translocations in lymphoid cancer. *Cell* 141, 27–38.
- Pasqualucci, L., Bhagat, G., Jankovic, M., Compagno, M., Smith, P., Muramatsu, M., Honjo, T., Morse, H.C., 3rd, Nussenzweig, M.C., and Dalla-Favera, R. (2008). AID is required for germinal center-derived lymphomagenesis. *Nat. Genet.* 40, 108–112.
- Pavri, R., Gazumyan, A., Jankovic, M., Di Virgilio, M., Klein, I., Ansarah-Sobrinho, C., Resch, W., Yamane, A., Reina San-Martin, B., Barreto, V., et al. (2010). Activation-induced cytidine deaminase targets DNA at sites of RNA polymerase II stalling by interaction with Spt5. *Cell* 143, 122–133.
- Pefanis, E., Wang, J., Rothschild, G., Lim, J., Chao, J., Rabadan, R., Economides, A.N., and Basu, U. (2014). Noncoding RNA transcription targets AID to divergently transcribed loci in B cells. *Nature* 514, 389–393.
- Perlot, T., Li, G., and Alt, F.W. (2008). Antisense transcripts from immunoglobulin heavy-chain locus V(D)J and switch regions. *Proc. Natl. Acad. Sci. USA* 105, 3843–3848.
- Petersen-Mahrt, S.K., Harris, R.S., and Neuberger, M.S. (2002). AID mutates *E. coli* suggesting a DNA deamination mechanism for antibody diversification. *Nature* 418, 99–103.

- Qian, J., Wang, Q., Dose, M., Pruett, N., Kieffer-Kwon, K.R., Resch, W., Liang, G., Tang, Z., Mathé, E., Benner, C., et al. (2014). B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity. *Cell* 159, 1524–1537.
- Revy, P., Muto, T., Levy, Y., Geissmann, F., Plebani, A., Sanal, O., Catalan, N., Forveille, M., Dufourcq-Labeledou, R., Gennery, A., et al. (2000). Activation-induced cytidine deaminase (AID) deficiency causes the autosomal recessive form of the Hyper-IgM syndrome (HIGM2). *Cell* 102, 565–575.
- Ruskin, B., and Green, M.R. (1985). An RNA processing activity that de-branches RNA lariats. *Science* 229, 135–140.
- Sen, D., and Gilbert, W. (1988). Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* 334, 364–366.
- Srisawat, C., and Engelke, D.R. (2001). Streptavidin aptamers: affinity tags for the study of RNAs and ribonucleoproteins. *RNA* 7, 632–641.
- Taylor, B.J., Wu, Y.L., and Rada, C. (2014). Active RNAP pre-initiation sites are highly mutated by cytidine deaminases in yeast, with AID targeting small RNA genes. *eLife* 3, e03553.
- Tsai, M.C., Manor, O., Wan, Y., Mosammaparast, N., Wang, J.K., Lan, F., Shi, Y., Segal, E., and Chang, H.Y. (2010). Long noncoding RNA as modular scaffold of histone modification complexes. *Science* 329, 689–693.
- Vaughn, J.P., Creacy, S.D., Routh, E.D., Joyner-Butt, C., Jenkins, G.S., Pauli, S., Nagamine, Y., and Akman, S.A. (2005). The DEXH protein product of the DHX36 gene is the major source of tetramolecular quadruplex G4-DNA resolving activity in HeLa cell lysates. *J. Biol. Chem.* 280, 38117–38120.
- Williamson, J.R., Raghuraman, M.K., and Cech, T.R. (1989). Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell* 59, 871–880.
- Xu, Y., Kaminaga, K., and Komiyama, M. (2008). Human telomeric RNA in G-quadruplex structure. *Nucleic Acids Symp. Ser. (Oxf.)*, 175–176.
- Xu, Z., Fulop, Z., Wu, G., Pone, E.J., Zhang, J., Mai, T., Thomas, L.M., Al-Qah-tani, A., White, C.A., Park, S.R., et al. (2010). 14-3-3 adaptor proteins recruit AID to 5'-AGCT-3'-rich switch regions for class switch recombination. *Nat. Struct. Mol. Biol.* 17, 1124–1135.
- Yamane, A., Resch, W., Kuo, N., Kuchen, S., Li, Z., Sun, H.W., Robbiani, D.F., McBride, K., Nussenzweig, M.C., and Casellas, R. (2011). Deep-sequencing identification of the genomic targets of the cytidine deaminase AID and its cofactor RPA in B lymphocytes. *Nat. Immunol.* 12, 62–69.
- Yu, K., and Lieber, M.R. (2003). Nucleic acid structures and enzymes in the immunoglobulin class switch recombination mechanism. *DNA Repair (Amst.)* 2, 1163–1174.
- Zalfa, F., Giorgi, M., Primerano, B., Moro, A., Di Penta, A., Reis, S., Oostra, B., and Bagni, C. (2003). The fragile X syndrome protein FMRP associates with BC1 RNA and regulates the translation of specific mRNAs at synapses. *Cell* 112, 317–327.
- Zarrin, A.A., Alt, F.W., Chaudhuri, J., Stokes, N., Kaushal, D., Du Pasquier, L., and Tian, M. (2004). An evolutionarily conserved target motif for immunoglobulin class-switch recombination. *Nat. Immunol.* 5, 1275–1281.
- Zhao, J., Sun, B.K., Erwin, J.A., Song, J.J., and Lee, J.T. (2008). Polycomb proteins targeted by a short repeat RNA to the mouse X chromosome. *Science* 322, 750–756.

RNA Exosome-Regulated Long Non-Coding RNA Transcription Controls Super-Enhancer Activity

Graphical Abstract



Authors

Evangelos Pefanis, Jiguang Wang, ..., Raul Rabadan, Uttiya Basu

Correspondence

rr2579@cumc.columbia.edu (R.R.),
ub2121@cumc.columbia.edu (U.B.)

In Brief

The RNA exosome regulates the expression of non-coding RNAs originating from enhancer regions, helping to coordinate their function with the activity of neighboring super-enhancers.

Highlights

- Identification of new lncRNAs and eRNAs from RNA exosome mutant transcriptomes
- Transcribed enhancer sequences are protected from genomic instability by RNA exosome
- Antisense RNAs substrates of the RNA exosome regulate super-enhancer activity
- eRNA-expressing element (lncRNA-CSR) promotes recombination in the IgH locus



RNA Exosome-Regulated Long Non-Coding RNA Transcription Controls Super-Enhancer Activity

Evangelos Pefanis,^{1,4,6} Jiguang Wang,^{2,6} Gerson Rothschild,^{1,6} Junghyun Lim,^{1,6} David Kazadi,¹ Jianbo Sun,¹ Alexander Federation,⁵ Jaime Chao,¹ Oliver Elliott,² Zhi-Ping Liu,³ Aris N. Economides,⁴ James E. Bradner,⁵ Raul Rabadan,^{2,*} and Uttiya Basu^{1,*}

¹Department of Microbiology and Immunology

²Department of Biomedical Informatics and Department of Systems Biology

College of Physicians and Surgeons, Columbia University, New York, NY 10032, USA

³Department of Biomedical Engineering, School of Control Science and Engineering, Shandong University, Jinan, Shandong 250061, China

⁴Regeneron Pharmaceuticals and Regeneron Genetics Center, Tarrytown, NY 10591, USA

⁵Dana-Farber Cancer Institute and Harvard Medical School, Boston, MA 02115, USA

⁶Co-first author

*Correspondence: rr2579@cumc.columbia.edu (R.R.), ub2121@cumc.columbia.edu (U.B.)

<http://dx.doi.org/10.1016/j.cell.2015.04.034>

SUMMARY

We have ablated the cellular RNA degradation machinery in differentiated B cells and pluripotent embryonic stem cells (ESCs) by conditional mutagenesis of core (*Exosc3*) and nuclear RNase (*Exosc10*) components of RNA exosome and identified a vast number of long non-coding RNAs (lncRNAs) and enhancer RNAs (eRNAs) with emergent functionality. Unexpectedly, eRNA-expressing regions accumulate R-loop structures upon RNA exosome ablation, thus demonstrating the role of RNA exosome in resolving deleterious DNA/RNA hybrids arising from active enhancers. We have uncovered a distal divergent eRNA-expressing element (lncRNA-CSR) engaged in long-range DNA interactions and regulating *IgH* 3' regulatory region super-enhancer function. CRISPR-Cas9-mediated ablation of lncRNA-CSR transcription decreases its chromosomal looping-mediated association with the *IgH* 3' regulatory region super-enhancer and leads to decreased class switch recombination efficiency. We propose that the RNA exosome protects divergently transcribed lncRNA expressing enhancers by resolving deleterious transcription-coupled secondary DNA structures, while also regulating long-range super-enhancer chromosomal interactions important for cellular function.

INTRODUCTION

Recent advances in RNA biology have revealed a plethora of non-coding RNA transcripts whose identity and functions were previously unknown. It has been postulated that transcription control of coding genes is modulated by non-coding RNAs such as enhancer RNAs (eRNAs) (Kim et al., 2010) and long intergenic non-coding RNAs (lincRNAs) (Rinn and Chang, 2012). Of

note, a significant number of non-coding RNAs are characterized as being expressed from regions proximal to the transcription start sites (TSSs) of coding genes. These transcripts include promoter-associated long RNAs (PALRs, >200 bp and bidirectional) (Kapranov et al., 2007), promoter-associated short RNAs (PASRs, 20–100 nt) (Kapranov et al., 2007), TSS-associated RNA (TSS-aRNA, small and divergently transcribed RNA) (Core et al., 2008; Seila et al., 2008), and transcription initiation RNAs (tiRNAs, 18 nt long and located 20 nt downstream of the coding TSS) (Taft et al., 2009). In addition, a large fraction of TSS-proximal transcriptional expenditure is dedicated to the production of unstable non-coding RNAs that are subject to RNA exosome-mediated degradation (PROMPTs, uaRNAs, xTSS-RNAs) (Flynn et al., 2011; Pefanis et al., 2014; Preker et al., 2008). Although the characteristics of these new RNA species may overlap, it is abundantly clear that these non-coding RNAs function in the regulation of transcription initiation and transcription elongation by various mechanisms, including control of RNA polII pausing and recruitment of chromatin modification factors (Flynn and Chang, 2012; Reyes-Turcu and Grewal, 2012; Shin et al., 2013).

Recently, some of these ncRNAs have been shown to be substrates of the RNA surveillance complex, RNA exosome (Anderson et al., 2014a, 2014b; Pefanis et al., 2014; Wan et al., 2012). The eukaryotic RNA exosome complex functions in both the nucleus and the cytoplasm. Nuclear exosome is involved in 3'-5' processing of rRNAs, sn/snoRNAs, degradation of hypomodified tRNAs, and cryptic unstable transcripts (CUTs), whereas cytoplasmic exosome is responsible for the degradation of aberrant mRNA species subject to nonsense mediated decay, non-stop decay, or no-go decay (Schmid and Jensen, 2008; Chlebowski et al., 2013). The eukaryotic exosome complex is composed of a nine subunit core, consisting of six distinct proteins forming a "ring" and three distinct RNA-binding-domain-containing proteins forming a "cap" structure required for the stabilization of the core structure. Enzymatic activity of the exosome complex is provided through two additional subunits: Rrp44 (*Dis3*) and Rrp6 (*Exosc10*) (Houseley et al., 2006; Januszynski and Lima, 2011; Liu et al., 2006; Lorentzen et al., 2008). Rrp6 is a nuclear-specific 3'-5' distributive exoribonuclease (Lykke-Andersen et al., 2009). Although in vitro Rrp6 and Dis3

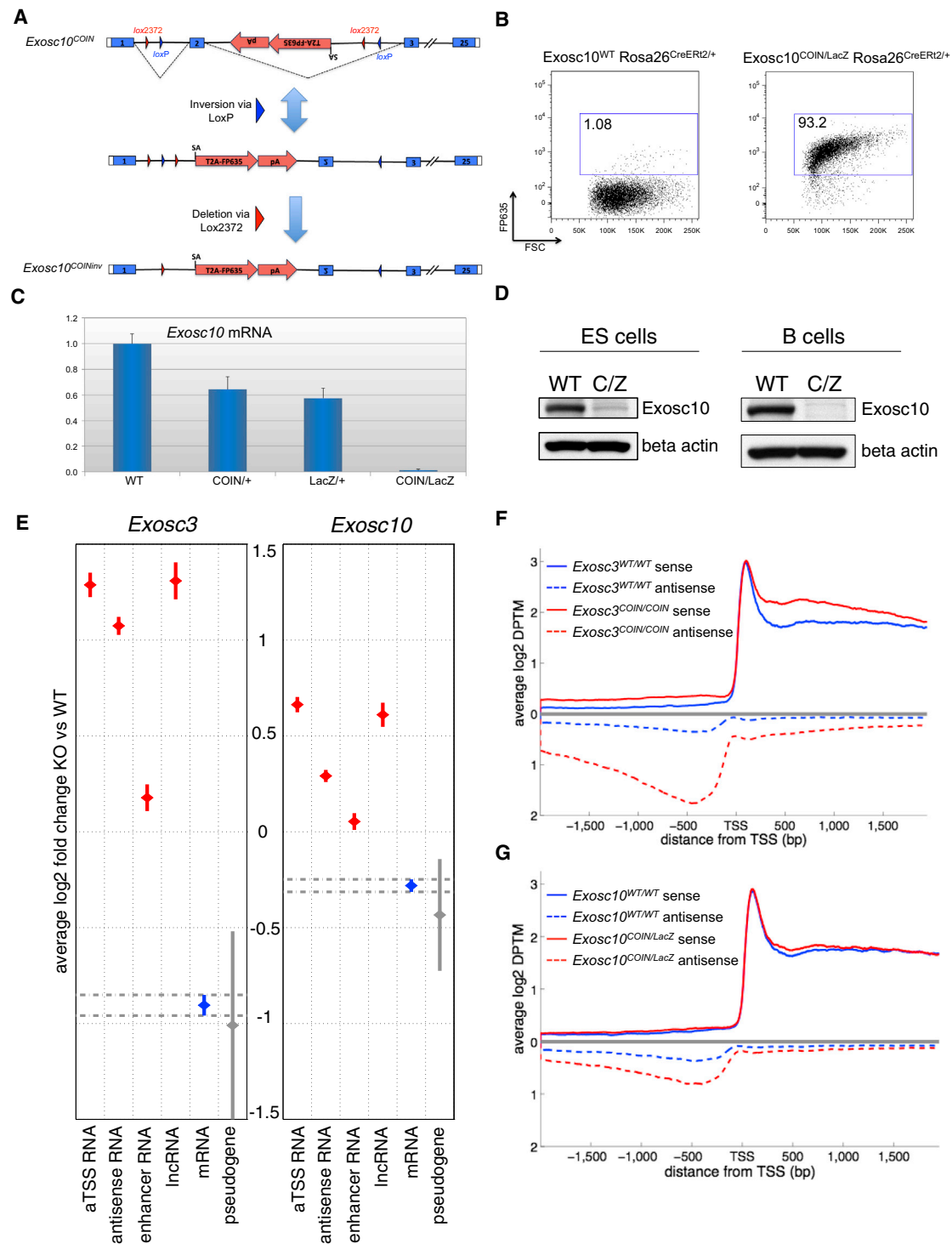


Figure 1. Generation of RNA Exosome Mutant ESCs and Transcriptome Analysis

(A) *Exosc10*^{COIN} allele and conversion to *Exosc10*^{COINinv}. Cre-mediated inversion of loxP pair (blue triangles) and subsequent deletion via lox2372 pair (red triangles). FP635-expressing terminal exon represented by red arrow. SA, splice acceptor.

(B) Induction of fluorescent reporter FP635 in *Exosc10*^{COIN/LacZ} *Rosa26*^{CreERT2/+} B cells following 4-OHT treatment.

(C) qRT-PCR analysis of *Exosc10* mRNA expression in 4-OHT-treated, LPS+IL-4-stimulated B cells. Indicated *Exosc10* genotypes on *Rosa26*^{CreERT2/+} background. Expression levels normalized to cyclophilin A (*Ppia*) and plotted relative to *Exosc10*^{WT/WT}. Splenic B cells were isolated and treated with 4-OHT for 24 hr, and the cells were then washed. Total cellular RNA was isolated after 72 hr of B cell culture. Three technical replicates; error bars represent SD.

(legend continued on next page)

bind the RNA exosome core (Exo9) independent of each other, Exo9 may interconnect the properties of the two RNase subunits in vivo (Schaeffer et al., 2009; Schaeffer and van Hoof, 2011; Wasmuth and Lima, 2012) so that different types of RNA substrates can be processed/degraded. Crystal structure analysis of an Rrp6-containing yeast RNA exosome complex suggests that Rrp6 may function in regulating the size of the central channel through which RNA traverses prior to degradation (Wasmuth et al., 2014). The true nature of Rrp6 function within the RNA exosome complex, via its distributive RNase activity and/or its contribution to central channel regulation, is incompletely understood. Moreover, mammalian RNA substrates of the RNA exosome complex with or without the Rrp6 component have not been systematically identified. The activity of the RNA exosome in co-transcriptionally degrading RNA plays a critical function in the nucleus, with recent observations in yeast and mammalian cells indicating a role for RNA degradation in early transcription termination (Colin et al., 2014; Hazelbaker et al., 2013; Lemay et al., 2014; Pefanis et al., 2014; Richard and Manley, 2009; Shah et al., 2014; Storb, 2014; Sun et al., 2013b). As such, the role of RNA exosome in chromatin-associated events is a major focus of ongoing research.

In this study, we reveal and analyze the transcriptomes of *Exosc3*- and *Exosc10*-ablated embryonic stem cells (ESCs) and B cells and identify a vast number of non-coding RNAs with emergent biological functionality. Strikingly, we find that the RNA exosome regulates the levels of divergently transcribed enhancer RNAs by promoting co-transcriptional silencing, thereby preventing the persistence of detrimental chromatin structures that can lead to genomic instability. Moreover, we provide evidence that RNA exosome substrate divergently transcribed loci may regulate interactions with super-enhancer loci. Thus, our study provides a mode of long-range chromatin regulation not previously described. As an example, we have identified the long non-coding RNA (lncRNA)-CSR-expressing locus and report its regulation of immunoglobulin heavy-chain DNA rearrangements by functionally interacting with the 3' regulatory region super-enhancer sequence (3'RR).

RESULTS

RNA Exosome Mutant ESCs and Mouse Models

To ascertain the role of the RNA exosome complex in the degradation of non-coding RNAs, we have generated mouse conditional alleles of *Exosc10* (expressing the distributive nuclease subunit Rrp6) (Figures S1A and S1B) and *Exosc3* (expressing the RNA exosome core subunit Rrp40) (Pefanis et al., 2014). Using these two approaches, inducible RNA exosome deficiency

was evaluated in either primary pluripotent embryonic stem cells or differentiated mature B cells. *Exosc10* and *Exosc3* allele schemes utilize Cre/lox conditional inversion (COIN) methodology to ablate normal gene expression upon exposure of the alleles to Cre recombinase activity (Economides et al., 2013; Pefanis et al., 2014). The salient feature of this approach, as utilized here, is the inversion of one or more endogenous coding exons resulting in the simultaneous "activation" of a fluorescent reporter terminal exon within the same locus (Figure 1A). *Exosc10*^{COIN/WT} mice were crossed with mice heterozygous for a null allele of *Exosc10* (*Exosc10*^{LacZ/WT}) to derive ESCs and B cells of the genotype *Exosc10*^{COIN/LacZ}. Similarly, we have generated *Exosc3*^{COIN/COIN} ESCs and B cells (Pefanis et al., 2014). Both *Exosc10*^{COIN/LacZ} and *Exosc3*^{COIN/COIN} cells also contain the inducible *ROSA26*^{CreERT2} allele allowing for rapid ablation of RNA exosome activity upon tamoxifen treatment. When B cells from *Exosc10*^{COIN/LacZ} mice were treated with 4-hydroxytamoxifen (4-OHT) ex vivo, inversion of the *Exosc10*^{COIN} allele was observed in more than 90% of the cells (Figure 1B). qRT-PCR assays performed on total cellular RNA demonstrated nearly complete loss of *Exosc10* mRNA in 4-OHT-treated *Exosc10*^{COIN/LacZ} B cells (Figure 1C). Western blotting of protein extracts from *Exosc10*^{COIN/LacZ} B cells and ESCs demonstrated severe loss of Rrp6 protein following 4-OHT, indicating robust ablation of *Exosc10* expression (Figure 1D). The RNA exosome previously has been implicated in catalyzing class switch recombination (CSR) in B cells by supporting the activity of activation-induced cytidine deaminase (AID) (Basu et al., 2011). Consistent with these observations, *Exosc10*-deficient B cells display reduced CSR efficiency as compared to wild-type (WT) littermate control B cells (Figure S1C) despite comparable expression of AID (Figure S1D). Finally, RNA-seq analysis of *Exosc10*^{COIN/LacZ} B cells and ESCs confirmed loss of *Exosc10* transcripts in both cell types (Figure S1E). Similarly, and consistent with previously published characterization of *Exosc3* ablation in *Exosc3*^{COIN/COIN} B cells, RNA-seq analysis demonstrated a clear loss of *Exosc3* transcripts in both *Exosc3*^{COIN/COIN} B cells and ESCs (Figure S1F).

Transcriptome of RNA Exosome Mutant ESCs and B Cells

We assembled the transcriptomes of littermate pairs of WT control and *Exosc10*^{COIN/LacZ} or *Exosc3*^{COIN/COIN} B cells and ESCs using next-generation RNA sequencing technology. The bioinformatics pipeline used for transcriptome reconstitution is outlined in Figure S2A and is described in further detail in the Extended Experimental Procedures. We find that, in the exotomes (exosome-deficient transcriptome) of *Exosc3*^{COIN/COIN}

(D) Western blot detection of Rrp6 (*Exosc10*) from *Exosc10*^{WT/WT} (WT) and *Exosc10*^{COIN/LacZ} (C/Z) protein extracts obtained from 4-OHT-treated ESCs and B cells. Indicated *Exosc10* genotypes on *ROSA26*^{CreERT2/+} background.

(E) Genome-wide differential expression level analysis of RNA subsets in *Exosc3* (left) and *Exosc10* (right) ablated mouse ESCs relative to WT littermate-matched ESCs. The error bars represent confidence interval of mean value estimated by an improved version of the Tukey-Kramer method (see Extended Experimental Procedures).

(F and G) Genome-wide TSS proximal expression profile in *Exosc3* (F) and *Exosc10* (G) ablated mouse ESCs. Sense and antisense transcript levels 2 kb flanking the TSS of annotated coding transcripts are indicated. ESCs were treated with 4-OHT for 24 hr and then further cultured for an additional 48 hr before total RNA isolation.

See also Figure S1 and Tables S1, S2, S3, and S4.

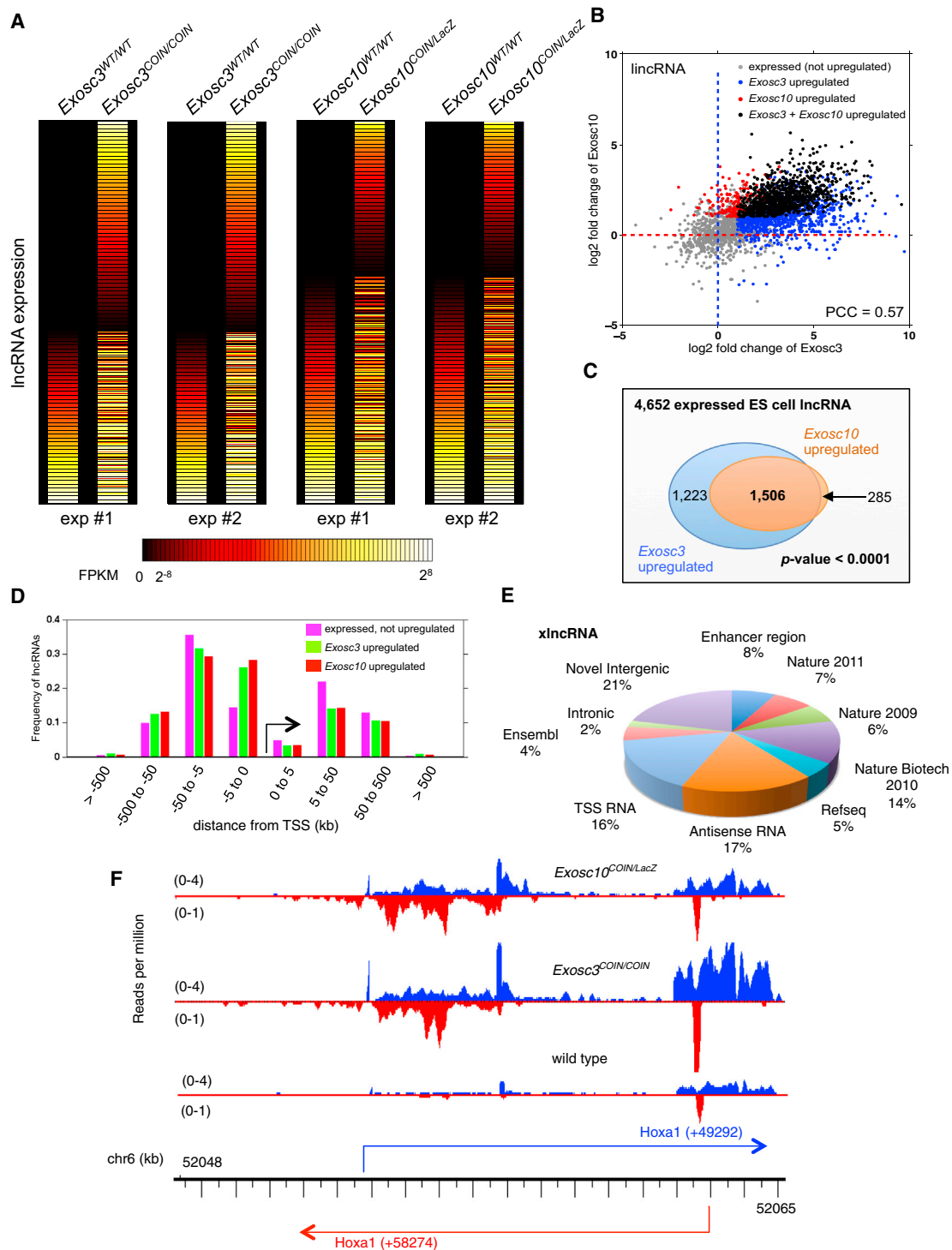


Figure 2. Identification and Characterization of RNA Exosome Targeted lncRNAs in ESCs

(A) Heatmap of lncRNAs expressed in *Exosc3*^{WT/WT}/*Exosc3*^{COIN/COIN} and *Exosc10*^{WT/WT}/*Exosc10*^{COIN/LacZ} genotype pairs. Horizontal lines represent different lncRNAs, which were ranked by their expression level in matched WT controls.

(B) Distribution of lncRNAs stabilized in the *Exosc3*-exotome (blue), *Exosc10*-exotome (red), and both *Exosc3* and *Exosc10* exotomes (black).

(C) Venn diagram demonstrating the distribution of overlapping lncRNAs in *Exosc3* (blue) and *Exosc10* (orange) exotomes.

(D) Distribution of x-lncRNA TSS distances from closest neighboring coding gene TSS genome wide.

(legend continued on next page)

(Figure 1E, left) and *Exosc10*^{COIN/LacZ} ESCs (Figure 1E, right), relative levels of lncRNAs, antisense RNAs, and eRNAs are significantly increased genome wide compared to WT control ESC transcriptomes. Comparing relative transcript accumulations of lncRNAs, antisense RNAs, and eRNAs indicates that these non-coding RNA subsets experience greater stabilization within the *Exosc3*^{COIN/COIN} exotome in comparison to the *Exosc10*^{COIN/LacZ} exotome genome wide. TSS antisense divergent RNAs are well-known substrates of the RNA exosome complex (Pefanis et al., 2014; Preker et al., 2008; Seila et al., 2008; Seila et al., 2009). Consistent with expectations, TSS-associated antisense RNAs are markedly stabilized within the *Exosc3*^{COIN/COIN} ESC transcriptome (Figure 1F). A list of antisense RNA in the body of the genes and around the genic TSS from B cell exotome and ESC exotome are provided in Tables S1 and S2, respectively. Relative to *Exosc3*-deficient cells, TSS-associated antisense transcripts are moderately stabilized within the *Exosc10*^{COIN/LacZ} ESC transcriptome (Figure 1G). Collectively, these results point toward a role for *Exosc10* in the degradation of a subset of RNA exosome-targeted lncRNAs (presumably fully represented via *Exosc3* ablation).

RNA Exosome Substrate Long Non-Coding RNA

Previously, it has been shown that enhancers express bidirectional, divergently transcribed, RNA exosome-sensitive, capped non-coding RNAs in human cell lines and primary mouse B cells (Andersson et al., 2014a, 2014b; Pefanis et al., 2014; Wan et al., 2012). Taking clues from these studies, we evaluated whether our RNA exosome mutant mouse models could be utilized for identifying eRNAs in pluripotent ESCs or lineage-committed matured B cells. Following the analysis pipeline described in the Extended Experimental Procedures, we observed that a subset of lncRNAs were strong substrates of RNA exosome. We describe such transcripts here as exosome substrate lncRNA (x-lncRNA). As shown via heatmap representation, both in *Exosc3*^{WT/WT}/*Exosc3*^{COIN/COIN} and in *Exosc10*^{WT/WT}/*Exosc10*^{COIN/LacZ} RNA-seq analysis pairs, multiple x-lncRNA loci are revealed in RNA exosome-deficient ESCs while weakly expressed in counterpart WT control cells (Figure 2A) (details of expression and genome coordinates of these transcripts supplied in Table S3). Next, we performed comparative expression analysis between *Exosc3* and *Exosc10* substrate x-lncRNAs and found that a significant number of, although not all, *Exosc3* x-lncRNAs also classify as *Exosc10* x-lncRNAs (Figure 2B). Specifically, of a total of 2,729 *Exosc3* x-lncRNAs in ESCs, 1,506 also fell within the cutoff for *Exosc10* x-lncRNAs (Figures 2C, S2B, and S2C; details in Table S3). Surprisingly, only 59% of *Exosc3* x-lncRNAs described here have been reported previously (Figures 2E and S2D). In fact, 236 of these identified x-lncRNAs are positioned close to enhancer sequences and thus may serve as RNA exosome target “x-eRNAs.”

Moreover, the accumulation of x-lncRNAs mostly maps within 5–50 kb from the TSS of known coding genes, making it possible that these lncRNAs regulate gene expression of distal genes via long-range chromatin interactions (Figure 2D). As indicated earlier, there are substantial numbers of lncRNAs that are quite unstably expressed in WT steady-state ESCs, but their identity cannot be confidently evaluated due to weak detection. However, RNA-seq analysis of *Exosc3*^{COIN/COIN} and/or *Exosc10*^{COIN/LacZ} cells provides a methodology for the detection and characterization of highly unstable lncRNA species. One such example is provided as the sense/antisense x-lncRNAs in the *Hoxa1* locus (Figure 2F). There are multiple species of antisense x-lncRNAs that are expressed in the *Hoxa1* locus (Figure 2F), whose detection is amplified in the *Exosc3*^{COIN/COIN} or *Exosc10*^{COIN/LacZ} exotomes.

RNA Exosome Substrate Enhancer RNA

Some enhancer RNAs (x-eRNAs) are predicted to form a subset of x-lncRNAs. Thus, we analyzed eRNA stability and identity in both *Exosc3* and *Exosc10* exotomes and found overlapping, as well as distinct, requirements for these two RNA exosome subunits (Figure 3A). All eRNAs that could be identified from ESCs are listed in Table S4. Of a total of 891 *Exosc3* x-eRNAs in ESCs, a subset of 423 displayed a significant enrichment with *Exosc10* loss (Figure 3B). In addition, 86% of the *Exosc3* x-eRNAs reported here are previously unrecognized. Of the 37 *Exosc3* x-eRNAs previously reported in VISTA, a subset of 18 was upregulated following *Exosc10* depletion (data not shown). In B cell exotomes, the degree of overlap between *Exosc3* and *Exosc10* x-eRNAs is reduced in comparison to ESC exotomes (Figure 3C). Of the 870 identified B cell *Exosc3* x-eRNAs, only 62 were *Exosc10* targets (Figure 3D). Representative *Exosc3* x-eRNAs within the *Cd83* locus were significantly upregulated in *Exosc3*^{COIN/COIN} B cells and modestly increased in *Exosc10*^{COIN/LacZ} B cells (Figure 3E).

x-lncRNA (or x-eRNA) expression is detectable in WT cells although significantly stabilized in *Exosc3*^{COIN/COIN} cells (Figure 3F). Moreover, in both B cells (Figure 3G) and ESCs (Figure 3H), the degree of conservation for x-lncRNAs genome wide is greater than a random control set of sequences, albeit lower in amount than protein-coding DNA sequences in the mouse genome. To determine the conservation of lncRNAs that we have identified in this study, we compared x-lncRNAs with human genes (genome version hg19) using the LiftOver tool (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>). The percentage of genes that are conserved between human and mouse is shown distributed with different cutoffs. In Figures 3G and 3H, equivalent numbers of coding genes/random genomic regions with similar length were generated as controls. For each group of genes, the percentage that is conserved between human and mouse (y axis) is calculated based on UCSC LiftOver tool

(E) The pie chart represents distribution of previously reported and newly identified lncRNAs from this study (Guttman et al., 2009, 2010, 2011). Each category represents RNAs unique to that category and non-overlapping with previous categories, with the initial category designated as “enhancer region lncRNAs” and proceeding clockwise.

(F) Expression of sense and antisense lncRNA expression profile at the *Hoxa1* locus identified from *Exosc3* and *Exosc10*-ablated ESCs. Based on RNA-seq read distribution, multiple lncRNAs are expressed in the sense and antisense directions.

See also Figure S2 and Table S3.

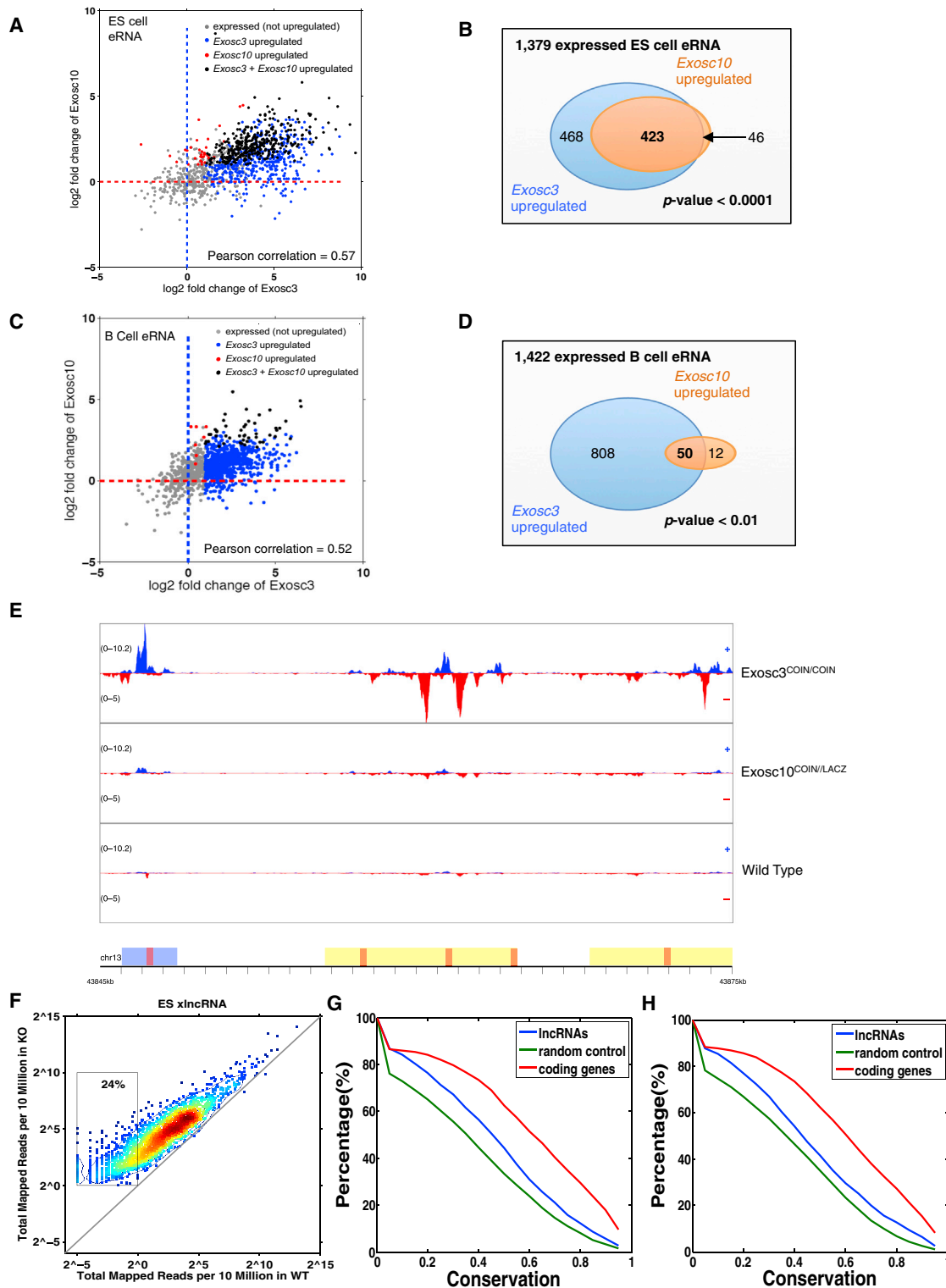


Figure 3. A Subset of Enhancer RNAs, x-eRNAs, Is an RNA Exosome Target in ESCs and B Cells

(A) Distribution of upregulated ESC x-eRNAs from *Exosc3* and *Exosc10* exotomes. Pearson correlation is indicated.
 (B) Overlap of identified x-eRNAs stabilized from degradation in ESC *Exosc3* (blue circle) and *Exosc10* (orange circle) exotomes.
 (C) Distribution of upregulated B cell x-eRNAs from *Exosc3* and *Exosc10* exotomes. Pearson correlation is indicated.
 (D) Overlap of identified x-eRNAs stabilized from degradation in B cell *Exosc3* (blue circle) and *Exosc10* (orange circle) exotomes.

(legend continued on next page)

with given cutoff (x axis) (details in [Extended Experimental Procedures](#)). Taking these observations into account, it is likely that many x-lncRNAs (and their subset x-eRNAs) are biologically functional. The dependency of the RNA exosome complex on Rrp6 (*Exosc10*) to degrade various subsets of ncRNAs may vary based on the type of ncRNA and/or the cell type. For example, xTSS-RNAs (one type of antisense RNA) in B cells ([Figure S3A](#)) or in ESCs ([Figure S3B](#)) have markedly increased representation in *Exosc3* exotomes in comparison to *Exosc10* exotomes. In contrast, antisense RNA levels arising from gene bodies were similar between *Exosc3* and *Exosc10* B cell ([Figure S3C](#)) and ESC exotomes ([Figure S3D](#)). Finally, to ascertain whether any major pathway was affected in the cells following RNA exosome activity depletion at the time points of RNA extraction, we performed gene set enrichment analysis (GSEA) in *Exosc3*^{WT/WT} and *Exosc3*^{COIN/COIN} ESCs. As would be expected, there were some perturbations in gene expression profiles in *Exosc3*^{COIN/COIN} ESCs, specifically gene sets related to organic acid transport and carboxylic acid transport (for details for GSEA of upregulated and downregulated pathways in *Exosc3*^{COIN/COIN} cells, see [Tables S5](#) and [S6](#), respectively).

RNA Exosome-Mediated Degradation of eRNAs Protects Cells from Genomic Instability by Preventing Formation of DNA/RNA Hybrids and by Promoting Heterochromatin Marks at Divergent Enhancers

Regions of the B cell genome beyond the *Ig* loci are susceptible to hypermutation due to AID activity and may then undergo chromosomal translocations involving *Ig* genes. Genomic loci susceptible to AID-induced chromosomal translocation break points may also accumulate x-eRNA reads in *Exosc3*^{COIN/COIN} B cells in comparison to *Exosc3*^{WT/WT} B cells. We observed that some *IgH* translocation partners identified through translocation capture techniques show x-eRNA expressing divergently transcribed enhancers as recurrent translocation hotspots. These include the *Birc3* enhancer ([Figure S4C](#)), as well as the *Nco3* enhancer ([Figure S4D](#)). These enhancer regions display overlapping sense and antisense RNA exosome substrate transcripts. Genomic overlaps between translocation breakpoints and x-eRNA-expressing regions provide evidence that RNA exosome-regulated enhancers in the B cell genome could be sensitive to DNA double-strand breaks resulting from AID, a physiologically expressed DNA mutator. Indeed, recently it has been ascertained that Rrp6 (*Exosc10*) plays a role in DNA double-strand break repair by affecting recruitment of ssDNA binding protein RPA ([Manfrini et al., 2015](#); [Marin-Vicente et al., 2015](#)). In fact, multiple studies indicate that AID-induced chro-

somal translocation sites in the B cell genome harbor RPA for DNA double-strand break repair ([Qian et al., 2014](#); [Yamane et al., 2013](#)).

Antisense RNAs that form co-transcriptional RNA/DNA hybrid structures called R-loops can initiate premature transcription termination and be a source of genomic instability ([Bhatia et al., 2014](#); [Pefanis et al., 2014](#); [Skourti-Stathaki et al., 2014](#)). In addition, such antisense RNAs can be substrates of the Dicer/Argonaute complex ([Skourti-Stathaki et al., 2014](#)) and RNA exosome ([Pefanis et al., 2014](#)). To investigate AID-independent DNA break formation in ESCs, we looked at whether x-eRNA-expressing regions are susceptible to genomic instability in RNA exosome-deficient cells due to formation of persistent R-loop structures. ESCs were irradiated with ionizing radiation (20 Gy) and allowed to recover over a period of 30 min. We evaluated three x-eRNA expressing loci neighboring *Klf6*, *Bcl6*, and *Cd38*. x-eRNA arising from these enhancer loci display divergent transcription and are sensitive to *Exosc3* function ([Figures S4E–S4G](#)). We evaluated the accumulation of DNA double-strand-break-associated γ -H2AX foci at divergent x-eRNA-expressing regions in *Exosc3*^{COIN/COIN} and *Exosc10*^{COIN/LacZ} cells. γ -H2AX accumulation at x-eRNA-expressing sequences was significantly enhanced in both *Exosc3* and *Exosc10* ablated ESC lines, implying a greater propensity for these sequences to undergo DNA double-strand breaks in the absence of functional RNA exosome complex ([Figure 4A](#)). Using immunoprecipitation assays with anti-DNA/RNA hybrid S9.6 antibody, we found that, in *Exosc3*^{COIN/COIN} and *Exosc10*^{COIN/LacZ} cells, x-eRNA-expressing regions are significantly enriched for RNase-H-sensitive DNA/RNA hybrid structures ([Figure 4B](#)). In contrast, an enhancer region in the ESC genome that does not demonstrate divergent transcription was not enriched for γ -H2AX foci or R-loops ([Figures S4I](#) and [S4J](#), respectively). These observations point toward the possibility that RNA exosome mutant ESCs are more prone to genomic instability insults at divergently transcribed enhancer sequences. Telomeric fluorescence in situ hybridization (FISH) assays performed on IR-treated *Exosc3*^{COIN/COIN} cells revealed a significantly greater frequency of chromosomal alteration in comparison to control *Exosc3*^{WT/WT} cells ([Figures S4A](#) and [S4B](#)). Taken together, RNA exosome-mediated degradation of RNA in DNA/RNA hybrids at divergently transcribed enhancer sequences might serve as a mechanism for the maintenance of genomic integrity in mammalian cells.

The established roles of H3K9me2 and HP1 γ chromatin marks in the cellular processes of chromatin condensation and transcriptional repression have recently been identified to appear at sites of transcription termination of antisense non-coding

(E) x-eRNA stabilization at annotated enhancers within the *Cd83* locus in *Exosc3*^{WT/WT}, *Exosc10*^{COIN/LacZ}, and *Exosc3*^{COIN/COIN} B cells. Sense and antisense RNA indicated in blue and red, respectively.

(F) Expression levels of x-lncRNAs in ES cells identified from *Exosc3*^{WT/WT} and *Exosc3*^{COIN/COIN} transcriptomes.

(G and H) Sequence conservation plot of coding genes (red), identified x-lncRNAs (blue), and random control (green) from B cells (G) and ESCs (H). To measure how conserved the lncRNAs we have identified are, we compared the lncRNAs with human genes (genome version hg19) by LiftOver tool (<https://genome.ucsc.edu/cgi-bin/hgLiftOver>). The percentage of genes that are conserved between human and mouse is shown according to different cutoffs. The same number of coding genes/random genomic regions with similar length are generated as controls. For each group of genes, the percentage of genes conserved between human and mouse (y axis) is calculated based on UCSC Liftover tool with given cutoff (x axis).

See also [Figure S3](#) and [Table S4](#).

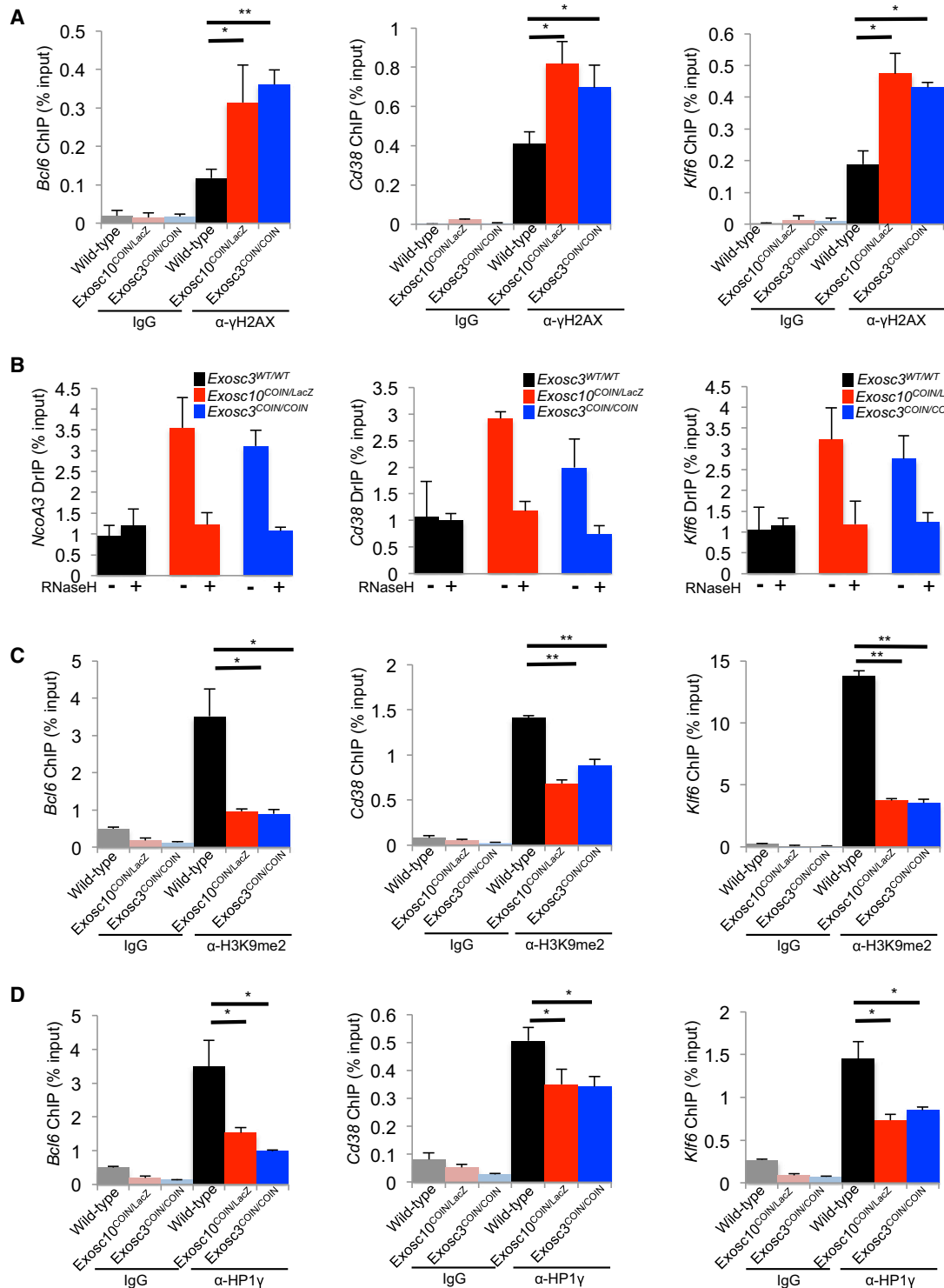


Figure 4. Genomic Instability in RNA Exosome-Deficient ESCs, along with Accumulation of DNA/RNA Hybrids and Loss of Chromatin-Silencing Markers H3K9me2 and HP1 γ at x-eRNA-Expressing Sequences

(A) γ H2AX immunoprecipitation for DNA double-strand breaks at enhancer sequences resident in the *Bcl6* (left), *Cd38* (middle), and *Klf6* (right) loci in WT, *Exosc10^{COIN/LacZ}*, and *Exosc3^{COIN/COIN}* ES cells following ionizing radiation treatment.

(B) DNA/RNA hybrid immunoprecipitation at *Nco3* (left), *Cd38* (middle), and *Klf6* (right) enhancers in WT, *Exosc10^{COIN/LacZ}*, and *Exosc3^{COIN/COIN}* ES cells.

(legend continued on next page)

RNAs (Skourti-Stathaki et al., 2014). Analysis of H3K9me2 (Figure 4C) and HP1 γ (Figure 4D) occupancy revealed decreased levels of these repressive chromatin marks at x-eRNA-expressing loci in *Exosc3*^{COIN/COIN} and *Exosc10*^{COIN/LacZ} cells. Thus, RNA exosome-mediated regulation of x-eRNA levels in cells could occur via two distinct mechanisms, namely via post-transcriptional RNA degradation or possibly through repression of RNA synthesis by promoting early transcription termination. In summary, we provide evidence that x-eRNA-expressing DNA sequences generate potentially deleterious DNA/RNA hybrids that might contribute to genomic instability.

x-eRNAs Have Biological Function at Super-Enhancer Sequences

Because enhancers are well-known modulators of gene expression, we evaluated x-eRNAs that arose from our analyses for functionality in controlling gene expression. We observed two peaks of sense and antisense transcription at regions upstream of the *Tgfb2* gene (Figure S5A). Using CRISPR-Cas9-mediated deletion of these lncRNA-expressing potential enhancer sequences in B cell line CH12F3, we observed a substantial decrease in the expression of *Tgfb2* mRNA by individually knocking out either of the two *Tgfb2* x-eRNA elements (Figure S5B).

We considered whether super-enhancer sequences, which are characterized by high density of individual enhancers and high regional enrichment for active chromatin marks, can generate RNA exosome substrate super-enhancer RNAs (x-seRNAs). As super-enhancer coordinates and functions can be identified in B cells using previously published bioinformatic pipelines (Lovén et al., 2013; Meng et al., 2014), we evaluated the expression of x-seRNAs in these cells. Our analysis revealed a significant enrichment of x-seRNAs in both *Exosc3* and *Exosc10* exotomes (Figure 5A). Relative to *Exosc3*^{COIN/COIN} cells, *Exosc10*-deficient cells retained significantly greater x-seRNA degradation activity, potentially due to RNA exosome complexes in these cells possessing the ability to utilize either the *Exosc10*-encoded Rrp6 or *Dis3*-encoded Rrp44 nuclease subunit in the degradation of x-seRNAs. We hypothesized that synthesis of antisense RNAs (either xTSS-RNA or those in the body of a gene) may functionally engage with super-enhancer elements to form higher-order chromosomal structures that may enable their local expression control. We sought such examples, i.e., super-enhancer sequences neighboring RNA exosome-sensitive antisense RNA (x-asRNA)-expressing genes and illustrate two examples here. First, a super-enhancer (Chr 10SE)-enhancer (overlapping the *Btg1* gene) pair separated by a distance of 232 kb from each other was found to express both x-seRNAs and xTSS-RNAs, respectively (Figure 5B). Accordingly, both the Chr 10SE x-seRNA and *Btg1* xTSS-RNA are contained within the *Exosc3* and *Exosc10* exotomes. As a second example, we identified a Chr1 SE that closely paired

with an x-asRNA arising within the *Btg2* locus. In this case, the separation of the SE and *Btg2* was a mere 4 kb, with both the x-seRNA and the x-asRNA being part of the *Exosc3* and *Exosc10* exotomes (Figure 5C). A statistical analysis of the proximity between xTSS-RNA-expressing genes and x-seRNA-expressing super-enhancer sequences illustrates a remarkable correlation that genes less than 310 kb from a SE are statistically far more likely to express antisense xTSS-RNAs ($p < 0.0001$; Figure 5D). The 310 kb distance between xTSS-RNA and x-seRNA-expressing sequences was set based on a genome-wide statistical analysis of distance between these elements in B cells. Beyond a distance of 310 kb from a super-enhancer, there is a consistent decrease in correlation of x-TSS-RNA expression (Figure S5C; details in Extended Experimental Procedures). These observations at individual loci such as *Btg1* and *Btg2*, along with genome-wide analyses, support a model whereby super-enhancer and counterpart gene interactions are controlled by expression and/or processing of RNA exosome substrate non-coding RNAs.

Molecular Evidence that Antisense RNA/Super-Enhancer RNA Expression Regulates Long-Range *IgH* Locus Recombination

A pair of divergently transcribed x-lncRNAs was found to be expressed at a 2.6 Mb distal region downstream of the 3'RR of the *IgH* locus. Both members of this x-lncRNA pair—named here as B930059L03Rik and lncRNA-CSR—were significantly more stable in *Exosc3*^{COIN/COIN} and *Exosc10*^{COIN/LacZ} B cells but also detectably expressed in WT control B cells (Figure 6C). A detailed map of this lncRNA-locus is shown in Figure S6A; no transcription factor binding sites were computationally predicted to overlap this region (Figure S6A). We proceeded to delete the lncRNA-CSR locus in CH12F3 cells using CRISPR-Cas9 and demonstrated complete loss of expression of lncRNA-CSR (Figure 6A). We found that lncRNA-CSR homozygous deleted CH12F3 cells expressed similar levels of the *IgH* locus recombination catalyst enzyme AID (Figure S5D). When lncRNA-CSR-deficient CH12F3 cells were assayed for CSR efficiency, they showed substantial defect for isotype switching to IgA (Figures 6B and S5E). Chromosome conformation capture (3C) (using lncRNA-CSR 3C primer Figure S6A and HS4 region primer Figure S6B) was performed to assess the interaction frequency of the lncRNA-CSR locus with regions of the *IgH* locus 3'RR super-enhancer (for details see Extended Experimental Procedures). Remarkably, we observed that the HS4 region of the *IgH* locus 3'RR interacts with the lncRNA-CSR locus. Deletion of the lncRNA-CSR sequence substantially decreased the interaction frequency between the deleted locus and the 3'RR HS4 region, whereas the canonical 3'RR and E μ interaction remained similar (Figure 6D). As can be seen from RNA-seq data, the antisense super-enhancer RNA peak corresponding to 3'RR HS4 (strongly visible in the *Exosc3*^{COIN/COIN} track) also corresponds

(C) Immunoprecipitation for heterochromatin marker H3K9me2 at enhancer sequences resident in the *Bcl6* (left), *Cd38* (middle), and *Klf6* (right) loci in WT, *Exosc10*^{COIN/LacZ}, and *Exosc3*^{COIN/COIN} ES cells.

(D) Immunoprecipitation for heterochromatin marker HP1 γ at enhancer sequences resident in the *Bcl6* (left), *Cd38* (middle), and *Klf6* (right) loci in WT, *Exosc10*^{COIN/LacZ}, and *Exosc3*^{COIN/COIN} ES cells. Each plot is a representation of three independent experiments performed (* $p < 0.05$; ** $p < 0.01$ by t test). See also Figure S4.

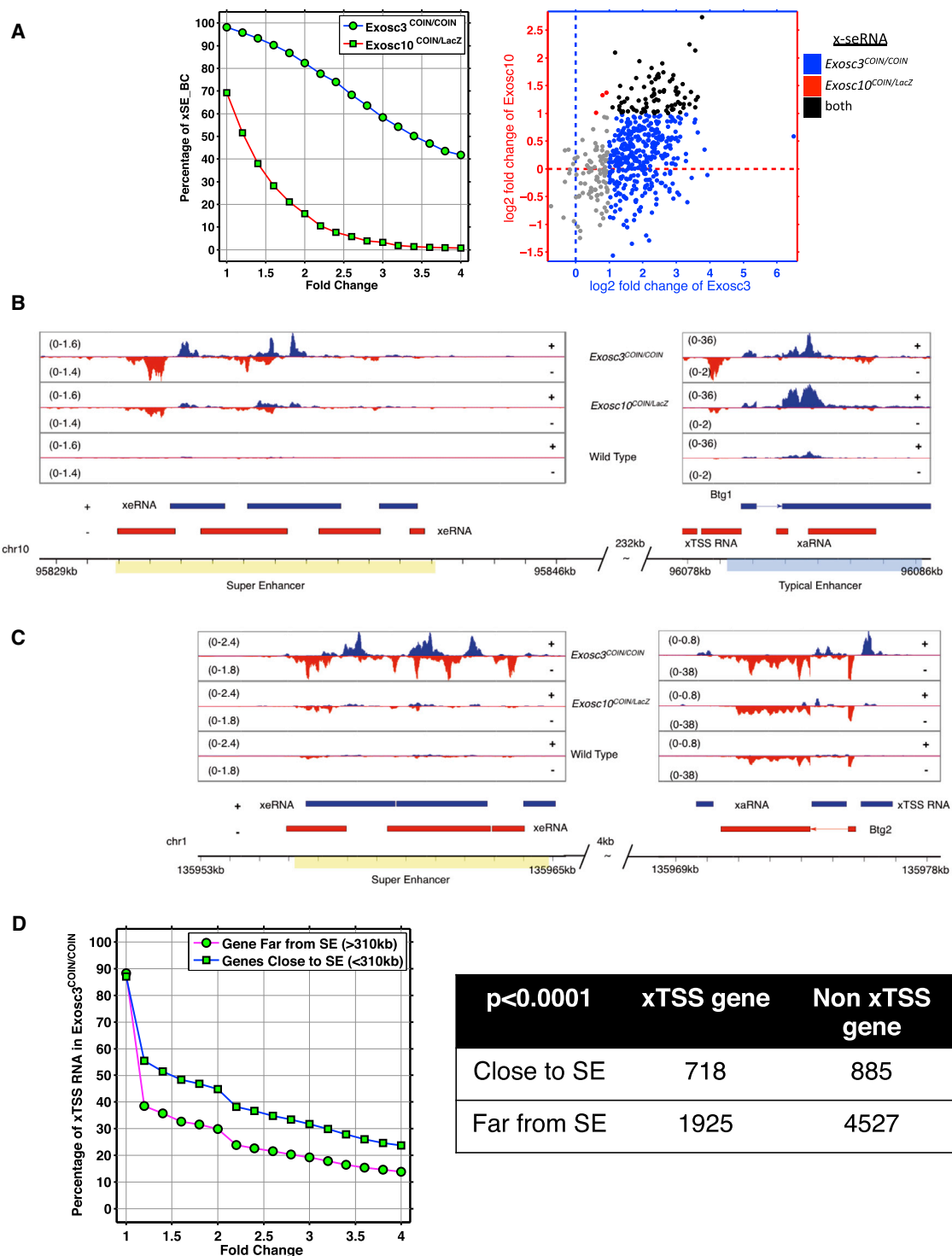


Figure 5. Super-Enhancer Sequences and Neighboring Conventional Enhancers or Coding Genes Express RNA Exosome Substrate Antisense RNAs

(A) Left: expression of super-enhancer RNAs at 529 annotated super-enhancers within *Exosc3* (*Exosc3*^{COIN/COIN}) and *Exosc10* (*Exosc10*^{COIN/LacZ}) exotomes. x axis indicates the cutoff of fold change, and y axis indicates the fraction of super-enhancers with higher expression (given x axis) compared with WT. Right: expression of *Exosc3* (blue), *Exosc10* (red), and overlapping (black) x-seRNAs in B cells.

(B) A super-enhancer resident in chromosome 10 (SEChr10) and neighboring conventional enhancer element resident at the *Btg1* locus express sense (blue) and antisense (red) x-eRNAs.

(legend continued on next page)

to the region of interaction with lncRNA-CSR based on DNA sequencing results from 3C assays (Figure 6C, bottom). The 3'RR HS4 region expresses multiple distinct x-seRNAs, as can be seen from the non-overlapping RNA-seq reads from the *Exosc3*^{COIN/COIN} transcriptome (Figure S6C). It is likely that the lncRNA-CSR element functions as a distal enhancer-like sequence and promotes the CSR-stimulating activity of the 3'RR super-enhancer via the interaction of the antisense lncRNA-CSR and the HS4 x-seRNA-expressing DNA regions. Thus, we provide functional evidence that RNA exosome substrate antisense RNA-expressing elements can interact with super-enhancer RNA-expressing regions to catalyze genomic rearrangement and organization.

We wanted to investigate the molecular mechanism of lncRNA-CSR transcription on the activity of 3'RR function in promoting CSR. The 3'RR is known to regulate transcription of switch region germline transcripts (GLTs) (Birshtein, 2014; Pinaud et al., 2011). IgS μ transcript levels were comparable between parental (WT) and Δ lncRNA-CSR CH12F3 clones (Figure 7A). On the other hand, we observed a significant suppression of IgA germline transcripts (IgS α) in the Δ lncRNA-CSR CH12F3 clones (Figure 7B). These observations point toward a role for lncRNA-CSR/HS4 interaction in regulating the transcription of downstream switch region transcripts at the S α locus. Whether this transcription regulation is similarly enforced at other switch regions can only be determined by generating mouse models deleted of the lncRNA-CSR locus. There is accumulation of long-range DNA rearrangements between the *IgH* (Klein et al., 2011) and lncRNA-CSR loci in B cells that overexpress AID (Figure S7A). Deletion of the lncRNA-CSR locus (Figure S6A) is presumed to disrupt its divergent transcription. We find, at least in these cells in which the transcription divergence is lost, H3K9me2 levels are decreased, raising the possibility that some level of heterochromatinization of these divergent sequences is important for their molecular activity to promote 3'RR interaction (Figure 7D). These observations are consistent with enhancer heterochromatinization regulation in ESCs by RNA exosome, as shown in Figures 4C and 4D. Finally, we evaluated the effect on 3'RR HS4-lncRNA-CSR interaction in B cells deficient in RNA exosome activity (*Exosc3*^{COIN/COIN}). We find that, in the absence of *Exosc3*, B cells have increased HS4-lncRNA-CSR interaction frequency relative to WT B cells (Figure 7C). However, increased interaction is not sufficient to promote CSR because the RNA exosome also regulates AID's DNA deamination activity in B cells (Basu et al., 2011; Pefanis et al., 2014; Sun et al., 2013a).

DISCUSSION

We envision that the identification of vast numbers of RNA exosome-targeted ncRNAs will enable the elucidation of their physiological roles in various developmental and gene expression regulatory pathways. Although many lncRNAs and their functions have been described (Bonasio and Shiekhattar, 2014; Rinn and

Chang, 2012; Sauvageau et al., 2013), our study identifies a subclass targeted by RNA exosome (x-lncRNA), many of which have not been reported previously. To explore, visualize, and analyze the landscape of these x-lncRNAs, we have generated a public browser showing strand-specific transcripts in the absence and presence of the RNA exosome complex subunits (see Extended Experimental Procedures). Such a tool may shed greater light on co-transcriptional processing dynamics at individual loci of interest and allow for generation of new hypotheses.

Recent findings have revealed the existence of vast numbers of intergenic and intragenic enhancer elements throughout the mammalian genome (Bonasio and Shiekhattar, 2014; Lam et al., 2014). How their activity is regulated is an exciting and open question. Enhancers generate eRNA transcripts whose biological role and regulation beyond chromatin remodeling are not well appreciated. In this study, we unravel the role of RNA exosome-mediated degradation of eRNAs expressed from divergently transcribed loci. We demonstrate that enhancer RNAs generate complexes with single-strand DNA that are protected from being converted to sites of genomic instability by the rapid action of the RNA exosome complex. The formation of R-looped DNA secondary structures can arise from failure to undergo proper transcriptional termination (Skourti-Stathaki et al., 2014). Early transcription termination serves as a mechanism for co-transcriptional RNA exosome recruitment (Lemay et al., 2014; Pefanis et al., 2014). Thus, in the absence of RNA exosome, x-eRNAs may accumulate not solely due to lack of RNA degradation but also due to failure of transiently forming R-loop structure-induced termination at enhancer loci (Skourti-Stathaki et al., 2014). Divergent transcription can create enhanced negative DNA supercoiling that, in turn, promotes the generation of ssDNA structures surrounding enhancer TSSs (Rhee and Pugh, 2012), thereby promoting DNA double-strand breaks and genomic instability (Pefanis et al., 2014). Such breaks could be caused by the activity of an endogenous DNA mutator such as cytidine deaminase AID or due to collisions of replication forks with stalled RNA polymerase complexes at these enhancer sequences (Kim and Jinks-Robertson, 2012). Sense/antisense x-eRNA pairs that form within the R-loop bubble may result in dsRNA that can be processed by RNAi factors, eventually leading to local accumulation of chromatin condensation marks such as H3K9me2 and HP1 γ (Skourti-Stathaki et al., 2014). Lack of RNA exosome activity may skew the ratio or abundance of sense and antisense eRNA transcripts, leading to impairment of RNAi pathway recruitment and heterochromatinization. Thus, RNA exosome may play an important role in promoting transcription termination-coupled silencing of divergent enhancer sequences genome wide.

Super-enhancers are large, densely packed enhancer elements that are occupied by master regulators of transcription and mediator proteins (Hnisz et al., 2013; Whyte et al., 2013). These elements are responsible for controlling transcription of diverse sets of tissue-specific gene expression programs. B cell

(C) A super-enhancer resident in chromosome 1 (SEChr1) and neighboring a conventional enhancer element resident at the *Btg2* locus express sense (blue) and antisense (red) x-eRNAs.

(D) Genome-wide correlation between proximity of super-enhancer location and antisense x-TSS-RNA expression at neighboring genes in B cells. See also Figure S5.

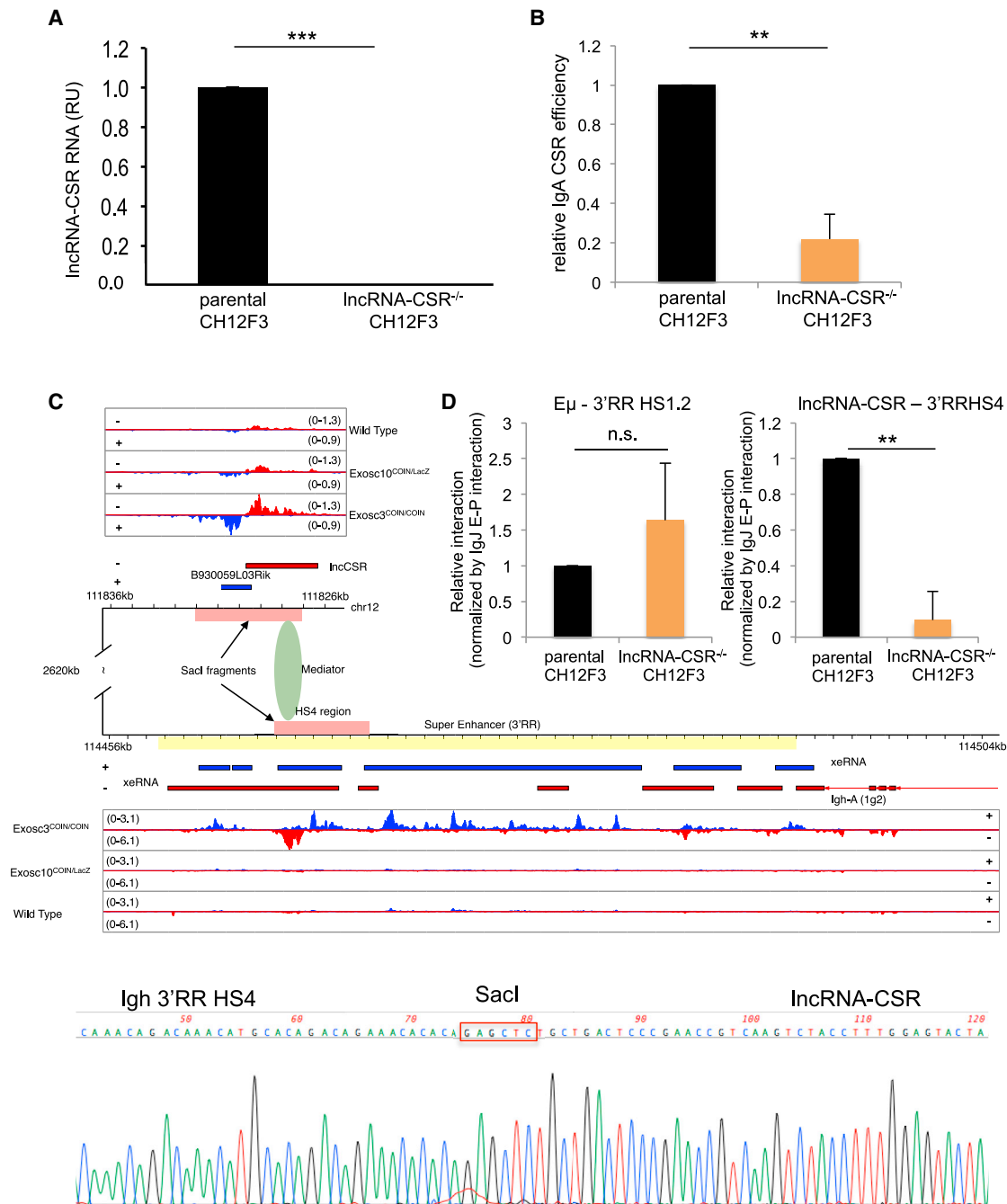


Figure 6. Identification of Divergently Expressed lncRNA-CSR at an Enhancer Region Controlling *IgH* Recombination in B Cells

(A) lncRNA-CSR expression in parental and lncRNA-CSR^{-/-} CH12F3 cells following CRISPR/Cas9 mediated deletion.

(B) IgA class switch recombination efficiency of lncRNA-CSR-deleted CH12F3 cells obtained from 18 independent lines of lncRNA-CSR^{-/-} CH12F3 cells.

(C) Top: expression profile of the lncRNA-CSR divergently transcribed enhancer locus that is stabilized in Exosc3^{COIN/COIN} and Exosc10^{COIN/LacZ} B cells. Middle: sense (blue) and antisense (red) tracks for 3' regulatory region super-enhancer transcription in Exosc3^{COIN/COIN} and Exosc10^{COIN/LacZ} B cells. Bottom: DNA sequencing of the 3C-derived joint PCR product of the super-enhancer *IgH* 3'RR HS4 sequence with the lncRNA-CSR enhancer sequence. The SacI site is contributed from the lncRNA-CSR locus and the HS4 locus and demonstrates the joining of the two pieces of DNA in the 3C assay.

(D) 3C assay determination of relative interaction frequency of Eμ with 3'RR HS1.2 and lncRNA-CSR locus with 3'RR HS4. **p < 0.01 and ***p < 0.001 by t test. See also Figures S5 and S6.

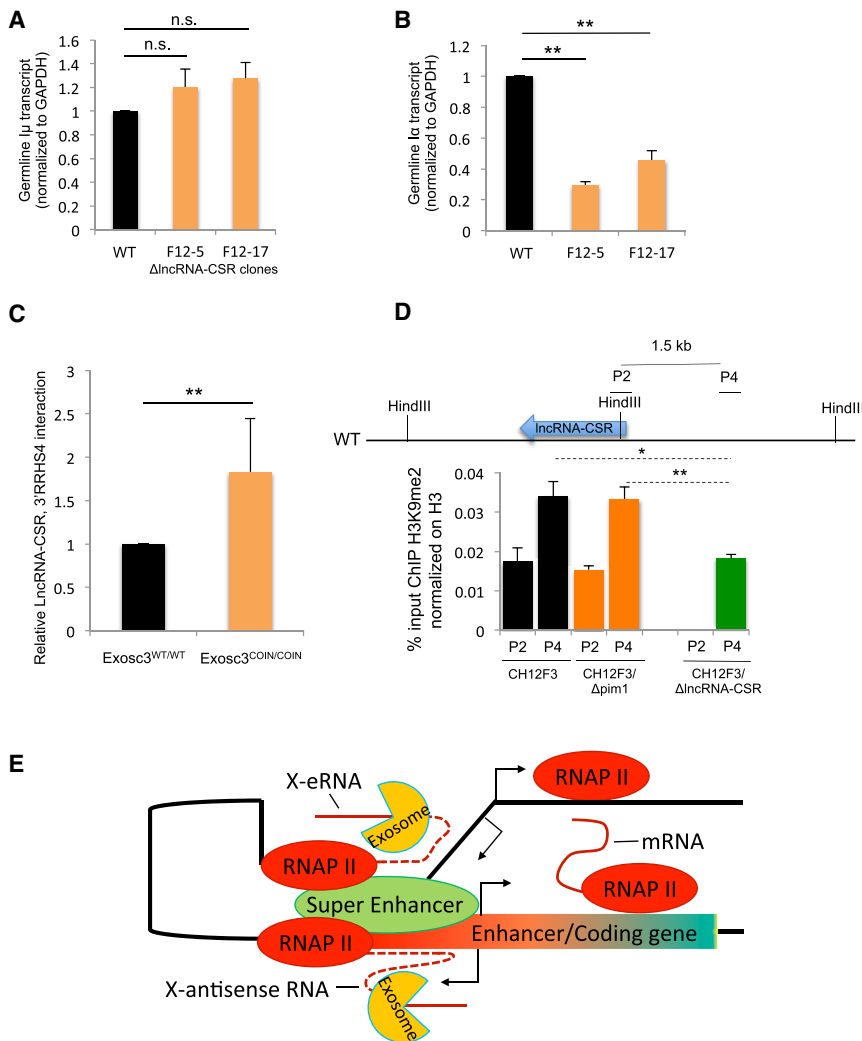


Figure 7. Mechanism of lncRNA-CSR-Mediated Suppression of 3'RR Super-Enhancer Function

(A) Germline transcripts at $l\mu$ in parental cells and two separate clones of lncRNA-CSR knockouts. Three independent sets of RNA were isolated for each cell line and assayed by qRT-PCR.

(B) Germline transcripts at $l\alpha$ in parental cells and two separate clones of lncRNA-CSR knockouts. Three independent sets of RNA were isolated for each cell line and assayed by qRT-PCR. ** $p < 0.01$ by t test.

(C) Chromosomal conformation capture performed on Exosc3^{WT/WT} and Exosc3^{COIN/COIN} B cells that were stimulated for CSR with LPS+IL4 for 24 hr. The frequency of 3'RRHS4 interaction with the lncRNA-CSR locus was measured by normalizing to an interaction downstream of the *Calr* gene locus. The experiment is a representation of three independently performed assays. ** $p < 0.01$ by t test.

(D) The accumulation of H3K9me2 marks (normalized to the presence of H3) in parental (CH12F3 cells), a random CRISPR/Cas9 mutated Δ pim1 (xTSS-RNA mutated) cell, and Δ lncRNA-CSR. Experiment is a representation of three independently performed assays. The ChIP assay primer pairs for various regions surrounding the lncRNA-CSR locus are shown in the top panel; * $p < 0.05$ and ** $p < 0.01$ by t test.

(E) A model of RNA exosome substrate x-seRNA-expressing super-enhancer interaction with the divergently transcribing promoter of another enhancer or protein coding gene. We postulate that the activity of RNA exosome to process the x-seRNA and x-eRNAs has a role in titrating the proper level of interaction between regulatory elements that ultimately control gene expression. See also Figure S7.

super-enhancers have been found to overlap large regions of the human genome susceptible to mutations in diffuse large B cell lymphomas (Chapuy et al., 2013) (Meng et al., 2014; Qian et al., 2014). We evaluated super-enhancers for the presence of RNA exosome-regulated transcripts and correspondingly identified x-seRNAs. Genes or canonical enhancers in proximity to super-enhancers express high levels of RNA exosome-regulated antisense RNAs around their TSSs (xTSS-RNAs) or within gene bodies (x-asRNAs). We hypothesize that super-enhancers may interact with genes under their regulation via mechanisms that depend upon transcription of RNA exosome-regulated transcripts. A test of this hypothesis was undertaken, and we observed that the divergently transcribed lncRNA-CSR enhancer element interacts with the HS4 region of the 3' regulatory region super-enhancer of the *IgH* locus to control class switch recombination. The dependence of a super-enhancer function on an interacting lncRNA-expressing divergent enhancer provides a newly identified mechanism of gene expression regulation (see Figure 7E for a proposed model). Whether the interaction is dependent upon direct RNA-protein complexes that are co-transcriptionally gener-

ated at the cognate pairs of enhancer/promoter and super-enhancer loci is a question of immediate interest. Furthermore, the observation that 3'RR x-seRNAs and lncRNA-CSR are substrates of RNA exosome provides the possibility that RNA exosome regulates long-distance genomic interactions either through its RNA degradation activities and/or through its ability to terminate transcription of ncRNAs at enhancers and super-enhancers.

EXPERIMENTAL PROCEDURES

Details of ChIP experiments, DNA/RNA hybrid immunoprecipitation, and 3C can be found in the Extended Experimental Procedures.

Exosc10^{COIN} Allele Design and Construction

A mouse *Exosc10* locus containing bacterial artificial chromosome (clone bMQ169f23) was modified using bacterial homologous recombination. Briefly, a *lox2372-loxP* array was inserted in the first intron of *Exosc10*. In a subsequent recombination event, an inverted *lox2372-loxP* array, inverted FP635 expressing terminal exon (COIN module) in antisense orientation to *Exosc10* transcription, and an FRT-flanked neo^r selection cassette were inserted within

a non-conserved region of *Exosc10* exon 2. The *Exosc10* COIN module contains a 3' splice acceptor sequence immediately followed by an in-frame T2A-FP635-pA cassette. *Exosc10^{COINneo}* BAC recombinants were screened by PCR across all four modified junctions and confirmed using restriction digestion and pulse field electrophoresis. A 20 kb fragment containing the entire *Exosc10^{COINneo}* modification was then subcloned into a plasmid containing a diphtheria toxin A (DTA) cassette. *Exosc10^{COINneo}* homology arms in the DTA vector were 6.7 and 8.2 kb. Linearized *Exosc10^{COINneo}* targeting vector was electroporated into ROSA26^{CreERT2/+}, 129S6/SvEv x C57BL/6 hybrid ESCs. Correctly targeted ESC clones were identified using external Southern blotting probes for both the upstream and downstream homology arms on HindIII or NsiI-digested genomic DNA, respectively. *Exosc10^{COIN/+}* chimeric mice were created via blastocyst injection of targeted ESCs. Mice with the greatest ESC-derived coat color contribution were crossed with Tg(ACTB:FLPe) mice to delete the neo^r selection cassette and germline transmit the *Exosc10^{COIN}* allele. The FLPe transgene was eliminated during backcrossing. All mouse experiments were conducted in accordance with approved Columbia University Institutional Animal Care and Use Committee protocols.

RNA-Seq Analysis

rRNA-depleted total RNA was prepared using the Ribo-Zero rRNA removal kit (Epicentre). Libraries were prepared with Illumina TruSeq and TruSeq Stranded total RNA sample prep kits and then sequenced with 50–60 million of 2 × 100 bp paired raw passing filters reads on an Illumina HiSeq 2000 V3 instrument at the Columbia Genome Center. The details of generation of exotomes from *Exosc3*-deficient or *Exosc10*-deficient B cells and ESCs and their subsequent analysis are described in the [Extended Experimental Procedures](#).

Transcriptome Reconstitution

Details of transcriptome reconstitution of the *Exosc3* and *Exosc10* exotomes from B cells and ESCs are described in detail in the [Extended Experimental Procedures](#), and the data are provided in [Tables S1](#), [S2](#), [S3](#), and [S4](#) and in the “Exotome browser,” which can be accessed from (<http://rabadan.c2b2.columbia.edu/cgi-bin/hgGateway>).

ACCESSION NUMBERS

The accession number for the RNA-sequencing data reported in this paper is Sequence Read Archive (SRA): SRP042355.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and six tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.04.034>.

AUTHOR CONTRIBUTIONS

E.P., J.W., and U.B. planned the studies, and E.P., J.W., G.R., R.R., and U.B. interpreted the data. Experiments were performed as follows: E.P. and J.C., undertook mouse model generation and RNA-seq studies; G.R. and J.L. performed 3C, ChIP, DRIP, and CRISPR/Cas9; J.W. conducted bioinformatic studies and designed all the pipelines for determining various RNA exosome substrate non-coding RNAs; A.N.E. advised on mouse model construct; R.R. oversaw bioinformatics and guided J.W.; D.K. and J.S. prepared and analyzed metaphases of RNA exosome mutant cells; J.W. and O.E. prepared exotome browser; A.F. and J.E.B. identified the super-enhancer sequences in B cells; and E.P., J.W., and U.B. wrote the manuscript, which was further refined by all of the other authors.

ACKNOWLEDGMENTS

We thank Christopher Lima (MSKCC, New York), Frederic Chedin (University of California, Davis), Saul Silverstein, Stephen Goff, Sankar Ghosh, Lorraine Symington, and members of the Basu lab for critical input and reagents. We thank Olivier Couronne of Columbia University genome center for

RNA sequencing and Victor Lin of Herbert Irving Cancer Center Transgenic facility for targeting and generation of *Exosc3^{COIN}* and *Exosc10^{COIN}* allele ESCs and mouse models. This work was supported by grants from NIH (1DP2D0008651-01) and NIAID (1R01AI099195-01A1) to U.B.; NIH (1R01CA185486-01; 1R01CA179044-01A1; 1U54CA121852-05) to R.R.; and F31AI098411-01A1 to D.K. U.B. is a scholar of the Irma Hirsch Charitable Trust and the Leukemia and Lymphoma Society.

Received: November 30, 2014

Revised: March 11, 2015

Accepted: April 20, 2015

Published: May 7, 2015

REFERENCES

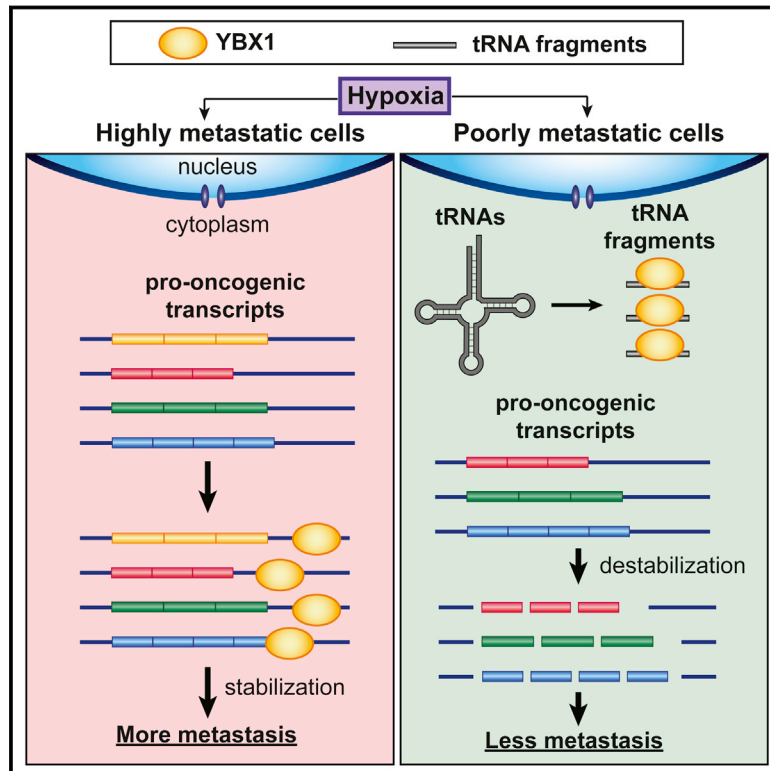
- Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmid, C., Suzuki, T., et al.; FANTOM Consortium (2014a). An atlas of active enhancers across human cell types and tissues. *Nature* **507**, 455–461.
- Andersson, R., Refsing Andersen, P., Valen, E., Core, L.J., Bornholdt, J., Boyd, M., Heick Jensen, T., and Sandelin, A. (2014b). Nuclear stability and transcriptional directionality separate functionally distinct RNA species. *Nat. Commun.* **5**, 5336.
- Basu, U., Meng, F.L., Keim, C., Grinstein, V., Pefanis, E., Eccleston, J., Zhang, T., Myers, D., Wasserman, C.R., Wesemann, D.R., et al. (2011). The RNA exosome targets the AID cytidine deaminase to both strands of transcribed duplex DNA substrates. *Cell* **144**, 353–363.
- Bhatia, V., Barroso, S.I., Garcia-Rubio, M.L., Tumini, E., Herrera-Moyano, E., and Aguilera, A. (2014). BRCA2 prevents R-loop accumulation and associates with TREX-2 mRNA export factor PCID2. *Nature* **511**, 362–365.
- Birshtein, B.K. (2014). Epigenetic Regulation of Individual Modules of the immunoglobulin heavy chain locus 3' Regulatory Region. *Front. Immunol.* **5**, 163.
- Bonasio, R., and Shiekhattar, R. (2014). Regulation of transcription by long noncoding RNAs. *Annu. Rev. Genet.* **48**, 433–455.
- Chapuy, B., McKeown, M.R., Lin, C.Y., Monti, S., Roemer, M.G., Qi, J., Rahl, P.B., Sun, H.H., Yeda, K.T., Doench, J.G., et al. (2013). Discovery and characterization of super-enhancer-associated dependencies in diffuse large B cell lymphoma. *Cancer Cell* **24**, 777–790.
- Chlebowski, A., Lubas, M., Jensen, T.H., and Dziembowski, A. (2013). RNA decay machines: the exosome. *Biochim. Biophys. Acta* **1829**, 552–560.
- Colin, J., Candelli, T., Porrua, O., Boulay, J., Zhu, C., Lacroute, F., Steinmetz, L.M., and Libri, D. (2014). Roadblock termination by reb1p restricts cryptic and readthrough transcription. *Mol. Cell* **56**, 667–680.
- Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science* **322**, 1845–1848.
- Economides, A.N., Frendewey, D., Yang, P., Dominguez, M.G., Dore, A.T., Lobov, I.B., Persaud, T., Rojas, J., McClain, J., Lengyel, P., et al. (2013). Conditionals by inversion provide a universal method for the generation of conditional alleles. *Proc. Natl. Acad. Sci. USA* **110**, E3179–E3188.
- Flynn, R.A., and Chang, H.Y. (2012). Active chromatin and noncoding RNAs: an intimate relationship. *Curr. Opin. Genet. Dev.* **22**, 172–178.
- Flynn, R.A., Almada, A.E., Zamudio, J.R., and Sharp, P.A. (2011). Antisense RNA polymerase II divergent transcripts are P-TEFb dependent and substrates for the RNA exosome. *Proc. Natl. Acad. Sci. USA* **108**, 10460–10465.
- Guttman, M., Amit, I., Garber, M., French, C., Lin, M.F., Feldser, D., Huarte, M., Zuk, O., Carey, B.W., Cassady, J.P., et al. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* **458**, 223–227.
- Guttman, M., Garber, M., Levin, J.Z., Donaghey, J., Robinson, J., Adiconis, X., Fan, L., Koziol, M.J., Gnirke, A., Nusbaum, C., et al. (2010). Ab initio

- reconstruction of cell type-specific transcriptomes in mouse reveals the conserved multi-exonic structure of lincRNAs. *Nat. Biotechnol.* 28, 503–510.
- Guttman, M., Donaghey, J., Carey, B.W., Garber, M., Grenier, J.K., Munson, G., Young, G., Lucas, A.B., Ach, R., Bruhn, L., et al. (2011). lincRNAs act in the circuitry controlling pluripotency and differentiation. *Nature* 477, 295–300.
- Hazebaker, D.Z., Marquardt, S., Wlotzka, W., and Buratowski, S. (2013). Kinetic competition between RNA Polymerase II and Sen1-dependent transcription termination. *Mol. Cell* 49, 55–66.
- Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. *Cell* 155, 934–947.
- Houseley, J., LaCava, J., and Tollervey, D. (2006). RNA-quality control by the exosome. *Nat. Rev. Mol. Cell Biol.* 7, 529–539.
- Januszky, K., and Lima, C.D. (2011). Structural components and architectures of RNA exosomes. *Adv. Exp. Med. Biol.* 702, 9–28.
- Kapranov, P., Cheng, J., Dike, S., Nix, D.A., Dutttagupta, R., Willingham, A.T., Stadler, P.F., Hertel, J., Hackermüller, J., Hofacker, I.L., et al. (2007). RNA maps reveal new RNA classes and a possible function for pervasive transcription. *Science* 316, 1484–1488.
- Kim, N., and Jinks-Robertson, S. (2012). Transcription as a source of genome instability. *Nat. Rev. Genet.* 13, 204–214.
- Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature* 465, 182–187.
- Klein, I.A., Resch, W., Jankovic, M., Oliveira, T., Yamane, A., Nakahashi, H., Di Virgilio, M., Bothmer, A., Nussenzweig, A., Robbiani, D.F., et al. (2011). Translocation-capture sequencing reveals the extent and nature of chromosomal rearrangements in B lymphocytes. *Cell* 147, 95–106.
- Lam, M.T., Li, W., Rosenfeld, M.G., and Glass, C.K. (2014). Enhancer RNAs and regulated transcriptional programs. *Trends Biochem. Sci.* 39, 170–182.
- Lemay, J.F., Larochelle, M., Marguerat, S., Atkinson, S., Bähler, J., and Bachand, F. (2014). The RNA exosome promotes transcription termination of backtracked RNA polymerase II. *Nat. Struct. Mol. Biol.* 21, 919–926.
- Liu, Q., Greimann, J.C., and Lima, C.D. (2006). Reconstitution, activities, and structure of the eukaryotic RNA exosome. *Cell* 127, 1223–1237.
- Lorentzen, E., Basquin, J., Tomecki, R., Dziembowski, A., and Conti, E. (2008). Structure of the active subunit of the yeast exosome core, Rrp44: diverse modes of substrate recruitment in the RNase II nuclease family. *Mol. Cell* 29, 717–728.
- Lovén, J., Hoke, H.A., Lin, C.Y., Lau, A., Orlando, D.A., Vakoc, C.R., Bradner, J.E., Lee, T.I., and Young, R.A. (2013). Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* 153, 320–334.
- Lykke-Andersen, S., Brodersen, D.E., and Jensen, T.H. (2009). Origins and activities of the eukaryotic exosome. *J. Cell Sci.* 122, 1487–1494.
- Manfrini, N., Trovesi, C., Wery, M., Martina, M., Cesena, D., Descrimes, M., Morillon, A., d'Adda di Fagnana, F., and Longhese, M.P. (2015). RNA-processing proteins regulate Mec1/ATR activation by promoting generation of RPA-coated ssDNA. *EMBO Rep.* 16, 221–231.
- Marin-Vicente, C., Domingo-Prim, J., Eberle, A.B., and Visa, N. (2015). RRP6/EXOSC10 is required for the repair of DNA double-strand breaks by homologous recombination. *J. Cell Sci.* 128, 1097–1107.
- Meng, F.L., Du, Z., Federation, A., Hu, J., Wang, Q., Kieffer-Kwon, K.R., Meyers, R.M., Amor, C., Wasserman, C.R., Neuberg, D., et al. (2014). Convergent transcription at intragenic super-enhancers targets AID-initiated genomic instability. *Cell* 159, 1538–1548.
- Pefanis, E., Wang, J., Rothschild, G., Lim, J., Chao, J., Rabadan, R., Economides, A.N., and Basu, U. (2014). Noncoding RNA transcription targets AID to divergently transcribed loci in B cells. *Nature* 514, 389–393.
- Pinaud, E., Marquet, M., Fiancette, R., Péron, S., Vincent-Fabert, C., Denizot, Y., and Cogné, M. (2011). The IgH locus 3' regulatory region: pulling the strings from behind. *Adv. Immunol.* 110, 27–70.
- Preker, P., Nielsen, J., Kammler, S., Lykke-Andersen, S., Christensen, M.S., Mapendano, C.K., Schierup, M.H., and Jensen, T.H. (2008). RNA exosome depletion reveals transcription upstream of active human promoters. *Science* 322, 1851–1854.
- Qian, J., Wang, Q., Dose, M., Pruett, N., Kieffer-Kwon, K.R., Resch, W., Liang, G., Tang, Z., Mathé, E., Benner, C., et al. (2014). B cell super-enhancers and regulatory clusters recruit AID tumorigenic activity. *Cell* 159, 1524–1537.
- Reyes-Turcu, F.E., and Grewal, S.I. (2012). Different means, same end-heterochromatin formation by RNAi and RNAi-independent RNA processing factors in fission yeast. *Curr. Opin. Genet. Dev.* 22, 156–163.
- Rhee, H.S., and Pugh, B.F. (2012). Genome-wide structure and organization of eukaryotic pre-initiation complexes. *Nature* 483, 295–301.
- Richard, P., and Manley, J.L. (2009). Transcription termination by nuclear RNA polymerases. *Genes Dev.* 23, 1247–1269.
- Rinn, J.L., and Chang, H.Y. (2012). Genome regulation by long noncoding RNAs. *Annu. Rev. Biochem.* 81, 145–166.
- Sauvageau, M., Goff, L.A., Lodato, S., Bonev, B., Groff, A.F., Gerhardinger, C., Sanchez-Gomez, D.B., Hacisuleyman, E., Li, E., Spence, M., et al. (2013). Multiple knockout mouse models reveal lincRNAs are required for life and brain development. *eLife* 2, e01749.
- Schaeffer, D., and van Hoof, A. (2011). Different nuclease requirements for exosome-mediated degradation of normal and nonstop mRNAs. *Proc. Natl. Acad. Sci. USA* 108, 2366–2371.
- Schaeffer, D., Tsanova, B., Barbas, A., Reis, F.P., Dastidar, E.G., Sanchez-Rotundo, M., Arraiano, C.M., and van Hoof, A. (2009). The exosome contains domains with specific endoribonuclease, exoribonuclease and cytoplasmic mRNA decay activities. *Nat. Struct. Mol. Biol.* 16, 56–62.
- Schmid, M., and Jensen, T.H. (2008). The exosome: a multipurpose RNA-decay machine. *Trends Biochem. Sci.* 33, 501–510.
- Seila, A.C., Calabrese, J.M., Levine, S.S., Yeo, G.W., Rahl, P.B., Flynn, R.A., Young, R.A., and Sharp, P.A. (2008). Divergent transcription from active promoters. *Science* 322, 1849–1851.
- Seila, A.C., Core, L.J., Lis, J.T., and Sharp, P.A. (2009). Divergent transcription: a new feature of active promoters. *Cell Cycle* 8, 2557–2564.
- Shah, S., Wittmann, S., Kilchert, C., and Vasiljeva, L. (2014). lncRNA recruits RNAi and the exosome to dynamically regulate *pho1* expression in response to phosphate levels in fission yeast. *Genes Dev.* 28, 231–244.
- Shin, J.H., Wang, H.L., Lee, J., Dinwiddie, B.L., Belostotsky, D.A., and Chekanova, J.A. (2013). The role of the Arabidopsis Exosome in siRNA-independent silencing of heterochromatic loci. *PLoS Genet.* 9, e1003411.
- Skourti-Stathaki, K., Kamieniarz-Gdula, K., and Proudfoot, N.J. (2014). R-loops induce repressive chromatin marks over mammalian gene terminators. *Nature* 516, 436–439.
- Storb, U. (2014). Why does somatic hypermutation by AID require transcription of its target genes? *Adv. Immunol.* 122, 253–277.
- Sun, J., Keim, C.D., Wang, J., Kazadi, D., Oliver, P.M., Rabadan, R., and Basu, U. (2013a). E3-ubiquitin ligase Nedd4 determines the fate of AID-associated RNA polymerase II in B cells. *Genes Dev.* 27, 1821–1833.
- Sun, J., Rothschild, G., Pefanis, E., and Basu, U. (2013b). Transcriptional stalling in B-lymphocytes: a mechanism for antibody diversification and maintenance of genomic integrity. *Transcription* 4, 127–135.
- Taft, R.J., Glazov, E.A., Cloonan, N., Simons, C., Stephen, S., Faulkner, G.J., Lassmann, T., Forrest, A.R., Grimmond, S.M., Schroder, K., et al. (2009). Tiny RNAs associated with transcription start sites in animals. *Nat. Genet.* 41, 572–578.
- Wan, Y., Qu, K., Ouyang, Z., Kertesz, M., Li, J., Tibshirani, R., Makino, D.L., Nutter, R.C., Segal, E., and Chang, H.Y. (2012). Genome-wide measurement of RNA folding energies. *Mol. Cell* 48, 169–181.

- Wasmuth, E.V., and Lima, C.D. (2012). Exo- and endoribonucleolytic activities of yeast cytoplasmic and nuclear RNA exosomes are dependent on the non-catalytic core and central channel. *Mol. Cell* 48, 133–144.
- Wasmuth, E.V., Januszyk, K., and Lima, C.D. (2014). Structure of an Rps6-RNA exosome complex bound to poly(A) RNA. *Nature* 511, 435–439.
- Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell* 153, 307–319.
- Yamane, A., Robbani, D.F., Resch, W., Bothmer, A., Nakahashi, H., Oliveira, T., Rommel, P.C., Brown, E.J., Nussenzweig, A., Nussenzweig, M.C., and Casellas, R. (2013). RPA accumulation during class switch recombination represents 5'-3' DNA-end resection during the S-G2/M phase of the cell cycle. *Cell Rep.* 3, 138–147.

Endogenous tRNA-Derived Fragments Suppress Breast Cancer Progression via YBX1 Displacement

Graphical Abstract



Authors

Hani Goodarzi, Xuhang Liu, ..., Lisa Fish, Sohail F. Tavazoie

Correspondence

stavazoie@mail.rockefeller.edu

In Brief

tRNA-derived fragments produced under hypoxic stress act as tumor suppressors through a post-transcriptional mechanism that leads to destabilization of many pro-oncogenic transcripts. Highly metastatic cells are capable of evading this mechanism by blunting the induction of tRFs during hypoxic conditions associated with cancer progression.

Highlights

- Hypoxic stress induces the production of tRNA-derived fragments (tRFs)
- This class of tRFs suppresses the development of breast cancer metastasis
- tRFs bind to oncogenic RNA-binding protein YBX1, displacing pro-oncogenic transcripts
- Highly metastatic cells blunt the induction of the tRFs during hypoxia

Accession Numbers

GSE63605



Endogenous tRNA-Derived Fragments Suppress Breast Cancer Progression via YBX1 Displacement

Hani Goodarzi,¹ Xuhang Liu,¹ Hoang C.B. Nguyen,¹ Steven Zhang,¹ Lisa Fish,¹ and Sohail F. Tavazoie^{1,*}

¹Laboratory of Systems Cancer Biology, Rockefeller University, New York, NY 10065, USA

*Correspondence: stavazoie@mail.rockefeller.edu

<http://dx.doi.org/10.1016/j.cell.2015.02.053>

SUMMARY

Upon exposure to stress, tRNAs are enzymatically cleaved, yielding distinct classes of tRNA-derived fragments (tRFs), yielding distinct classes of tRFs. We identify a novel class of tRFs derived from tRNA^{Glu}, tRNA^{Asp}, tRNA^{Gly}, and tRNA^{Tyr} that, upon induction, suppress the stability of multiple oncogenic transcripts in breast cancer cells by displacing their 3' untranslated regions (UTRs) from the RNA-binding protein YBX1. This mode of post-transcriptional silencing is sequence specific, as these fragments all share a common motif that matches the YBX1 recognition sequence. Loss-of-function and gain-of-function studies, using anti-sense locked-nucleic acids (LNAs) and synthetic RNA mimetics, respectively, revealed that these fragments suppress growth under serum-starvation, cancer cell invasion, and metastasis by breast cancer cells. Highly metastatic cells evade this tumor-suppressive pathway by attenuating the induction of these tRFs. Our findings reveal a tumor-suppressive role for specific tRNA-derived fragments and describe a molecular mechanism for their action. This transcript displacement-based mechanism may generalize to other tRNA, ribosomal-RNA, and sno-RNA fragments.

INTRODUCTION

Transfer RNA-derived RNA fragments (tRFs) belong to a family of short non-coding RNAs (ncRNAs) present in most organisms. These RNAs can be both constitutively generated and produced in the context of stress. Constitutive tRFs are thought to arise from ribonucleolytic processing of tRNAs by Dicer (Cole et al., 2009) and RNase Z (Lee et al., 2009). The generation of stress-induced tRFs, also known as stress-induced fragments (tiRNAs), has been shown to occur via the action of specific ribonucleases such as Angiogenin (Fu et al., 2009). Although tRNAs are one of the most abundant ncRNA molecules in the cell (~10% of total cellular RNA), only a small fraction of tRNAs are cleaved to produce tRFs (Thompson and Parker, 2009). Multiple classes of tRFs have been identified in various cell types and organisms and induced by various conditions. These classes are defined by the position of the tRNA cleavage site that gives rise to

tRFs, and these classes include 5'- and 3'-tRNA halves (cleaved in the anti-codon loop), 5'- and 3'-tRFs (also known as 3'CCA tRF), and 3'U tRFs, among others (Gebetsberger and Polacek, 2013).

Stress-induced tRFs have been reported to mediate a stress response, which results in stress granule assembly and inhibition of protein synthesis (Emara et al., 2010). Moreover, these tRFs can impact a number of cellular functions, such as cell proliferation and mediating RNA inactivation through Argonaute engagement (Gebetsberger and Polacek, 2013). In this study, we sought to investigate whether tRFs could play a role in metastatic progression. We reasoned that tRFs could have roles in cancer progression analogous to those of specific microRNAs (Krol et al., 2010). We also reasoned that because hypoxia is a major stress encountered by cells during cancer progression, tRFs induced under hypoxic conditions may act to curb metastatic progression. By employing next-generation small-RNA (smRNA) sequencing, we identified a group of tRFs that were upregulated under hypoxia in breast cancer cells as well as in non-transformed mammary epithelial cells. Interestingly, highly metastatic breast cancer cells did not display induction of these tRFs under hypoxia, suggesting a potential role for these molecules in cancer progression. We identified a common sequence motif present in these hypoxia-induced fragments, suggesting they may interact with a common *trans* factor. By using one of these tRFs (tRF^{Glu}) as bait, we immunoprecipitated and identified the RNA-binding protein YBX1 as a *trans* factor whose mRNA-stabilizing activity is repressed by these fragments.

YBX1 is a versatile RNA-binding protein with a variety of interacting partners. It is involved in many key cellular pathways, and its genetic inactivation leads to embryonic lethality (Uchiumi et al., 2006). Importantly, it is highly overexpressed in multiple cancer types (Jürchott et al., 2010; Matsumoto and Bay, 2005; Wu et al., 2012). By combining molecular, biochemical, and computational approaches, we find that tRFs bind YBX1 and displace a number of known oncogenic transcripts from YBX1, thereby antagonizing YBX1 activity. YBX1 stabilizes these oncogenic transcripts and mediates their enhanced expression. The displacement of these oncogenic transcripts by tRFs represses their stability and expression—thereby suppressing metastatic progression.

RESULTS

Systematic Identification of tRFs in Breast Cancer Cells

Tumor cells encounter various cellular stresses during the course of cancer progression. A critical stress is reduced access to

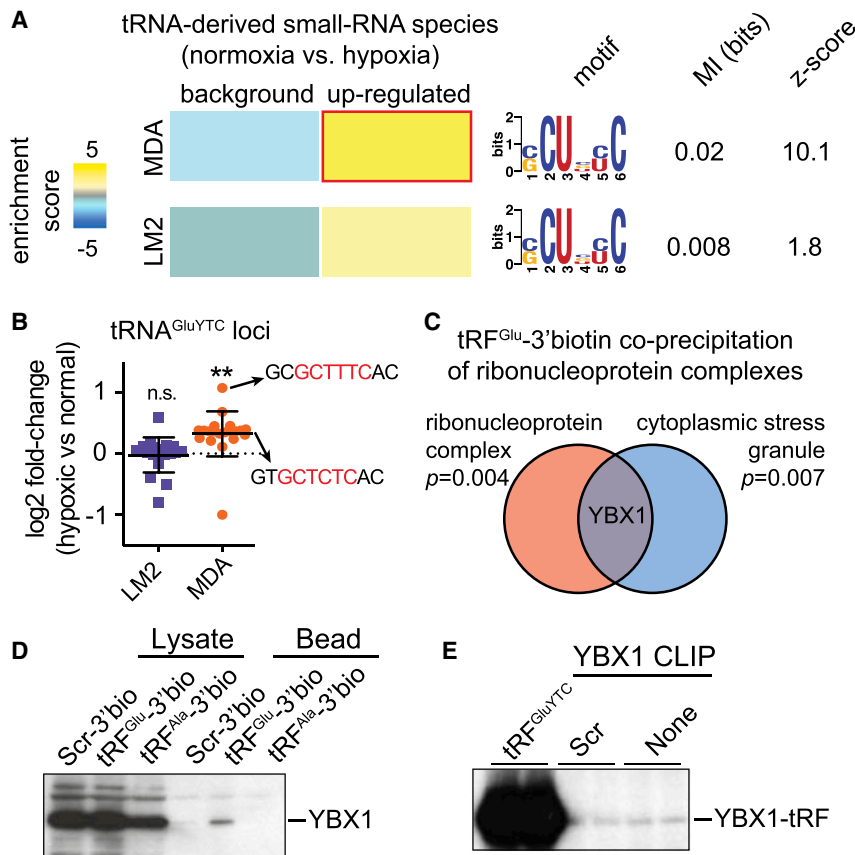


Figure 1. Genome-wide Profiling of tRFs in Breast Cancer MDA-231 Cells under Normal and Hypoxic Conditions

(A) The linear motif SCUBYC was enriched in RNA fragments mapping to tRNA loci that were upregulated in MDA-parental cells, but not MDA-LM2 cells, under hypoxic conditions. Shown are the mutual information values and their associated Z scores for the discovered motif in both cell lines (Elemento et al., 2007). The enrichment score (positive for enrichment and negative for depletion), presented as logP (hypergeometric p value), is also shown as a heatmap with blue showing depletion and yellow showing enrichment of the SCUBYC motif among the sequences in each cluster. The red border marks statistical significance of the enrichment score.

(B) The levels of tRFs derived from tRNA^{Glu} were significantly enhanced under hypoxic conditions in MDA-parental cells but not in MDA-LM2 cells. The log fold-change was calculated from the smRNA sequencing data. The p value was calculated using Wilcoxon rank-sum test. Two exemplary tRFs that contain the SCUBYC motif are also indicated.

(C) Streptavidin beads were used to co-precipitate proteins interacting with a 3'-biotinylated synthetic tRF^{Glu} mimetic and scrambled oligonucleotide in vivo. YBX1 was identified as a potential interacting partner based on the identity of the annotated RNA-protein complexes enriched among the tRF^{Glu} co-precipitated RNA-binding proteins.

(D) 3'-biotinylated synthetic oligonucleotides were used to co-precipitate YBX1. In addition to the scrambled RNA, a tRNA^{Ala}-derived fragment, which does not carry the identified motif, was also included as control. Western blotting was performed to detect YBX1 in the eluate from each sample.

(E) MDA-LM2 cells were transfected with a 21 nt synthetic tRF^{Glu} mimetic (unlabeled), also shown are scrambled transfected mimetic and untransfected cells as controls. After crosslinking immunoprecipitation of endogenous YBX1, and radiolabeling of the RNA population, a strong interaction between the transfected tRF^{Glu} mimetic and endogenous YBX1 was observed.

Error bars in all panels indicate SEM unless otherwise specified.

oxygen, a condition known as hypoxia (Moyer, 2012; Wilson and Hay, 2011). Multiple regulatory programs are co-opted by tumor cells to counteract the negative impacts of hypoxic stress (Bristow and Hill, 2008). For example, the stabilization and activation of the transcription factor HIF1 α under hypoxia results in the activation of vascular endothelial growth factor (VEGF, angiogenesis; Shen and Kaelin, 2013), GLUT1 (glucose transport), and carbonic anhydrase IX (CA9, pH regulation; Semenza, 1999). Recently, it was reported that tRFs are produced under hypoxia and during other stress conditions (Fu et al., 2009). Given the ability of hypoxia to significantly modulate the regulatory landscape of the cell at both transcriptional and post-transcriptional levels, we searched for tRNA fragments with potential regulatory roles that are modulated under hypoxic conditions in cancer cells. To do so, we performed smRNA sequencing of breast cancer cells (MDA-MB-231, hereafter termed MDA-parental). We observed that a sizeable fraction (~4%) of the smRNA population originated from tRNAs and therefore could be categorized as tRFs. We observed > 10 smRNA reads mapping to each of more than 300 tRNA loci across these samples.

tRFs belong to a class of smRNAs that are generated through endonucleolytic cleavage of tRNAs (Gebetsberger and Polacek,

2013; Thompson and Parker, 2009). These fragments have been detected in bacteria, yeast, and mammalian cells under normal and stress conditions (Gebetsberger and Polacek, 2013; Lee et al., 2009). Although tRNA fragments were first detected in the urine of cancer patients more than three decades ago, and at the time were proposed to be oncogenic molecules (Borek et al., 1977; Speer et al., 1979), their roles and mechanisms of action during cancer progression remain uncharacterized.

Consistent with their potential roles in stress response, tRF levels in breast cancer cells significantly increased under hypoxia ($p < 1e-6$, Figure S1A). Interestingly, the induction of these fragments by hypoxia was significantly blunted in MDA-LM2 cells—a highly metastatic sub-population derived through in vivo selection from the MDA-parental population (Figure S1A; Minn et al., 2005). These findings suggested that highly metastatic cells evade the upregulation of these fragments under hypoxic conditions. Sequence analysis of the tRNAs with hypoxia-induced fragmentation revealed the significant enrichment of a common linear sequence motif (SCUBYC; Figure 1A). This motif was not significantly enriched in the tRNA loci whose fragments were upregulated in highly metastatic MDA-LM2 cells, further highlighting the absence of a concerted upregulation of these

tRFs in highly metastatic cells (Figures 1A and S1A). The identification of this sequence motif among hypoxia-induced tRFs in MDA-parental cells raised the possibility that this element serves as a binding site for a common *trans* factor that potentially interacts with these small ncRNAs in vivo. For example, among tRNA^{Glu}-derived fragments, which were significantly upregulated under hypoxia in MDA-parental but not MDA-LM2 cells, several tRF species carried instances of the SCUBYC sequence motif described above (Figure 1B). In order to identify the unknown *trans* factor that may recognize this sequence motif, we used synthetic oligonucleotides from tRNA^{Glu} as bait in an in vitro co-precipitation experiment. A 21 nt 3'-biotinylated oligonucleotide carrying an instance of the identified motif was immobilized on streptavidin beads and was subsequently used to co-precipitate the interacting protein complexes. A scrambled RNA was processed in parallel as the control to measure the co-precipitation of proteins above background. In-solution digestion followed by mass-spectrometry was employed to determine the identity of the co-precipitated proteins. Gene-set enrichment analysis revealed that proteins annotated as components of "ribonucleoprotein complex" and "stress granule complex" were significantly over-represented among co-precipitated proteins (Figure 1C). YBX1, which showed 5-fold enrichment above background (Figure S1B), is the only protein that shares both of these annotations (Figure 1C); as such, we chose to further study this RNA-binding protein as a candidate tRF-interacting protein. To validate this interaction, we performed reciprocal co-immunoprecipitations and detected binding of YBX1 to an exogenously transfected 3'-biotinylated tRF^{Glu} mimetic, but not to the scrambled RNA or tRF^{Ala} controls (Figure 1D). Consistently, we also detected the binding of endogenous YBX1 to the tRF^{Glu} mimetic, but not the scrambled RNA control (Figure 1E).

Specific tRFs Interact with YBX1

YBX1, a multifunctional RNA-binding protein and a member of the Y box-binding protein family, has been implicated in various aspects of RNA biology. Importantly, it is a known modulator of RNA translation and stability and has been implicated in tRNA^{Ala}-mediated inhibition of ribosomal activity in vivo (Ivanov et al., 2011). Consistently, while resolving YBX1-crosslinked RNAs on a polyacrylamide gel, we observed a prominent smRNA band in addition to longer RNA species interacting with YBX1 (Figures 1E and S1C). We hypothesized that in addition to tRF^{Glu}, which was used as bait to identify YBX1, YBX1 may also interact with a broader population of endogenous smRNAs in the cell. To generate a detailed and precise snapshot of genome-wide YBX1-RNA interactions across long- and short RNA species, we performed crosslinking immunoprecipitation of endogenous YBX1 followed by high-throughput sequencing (CLIP-seq) in human MDA-parental breast cancer cells. Analysis of the CLIP-seq data provided us with the first in vivo YBX1 transcript interactome revealing more than 4,000 endogenous transcripts bound by YBX1. The majority of YBX1-binding sites were localized to 3' untranslated regions (3' UTRs) and exons, whereas minimal binding was detected in 5' UTRs and in intronic sequences (Figure 2A). A large number of cellular processes and pathways, including RNA processing, translation, cell cycle, glucose catabolism, spindle organization, and additional key signaling and

stress-response pathways, were over-represented among the YBX1-bound transcripts (Figure S1D). The breadth and diversity of the YBX1 regulon (Figures S1E and S1F) highlight its role in RNA homeostasis and growth, as well as its importance for cellular response to internal and external stimuli.

Consistent with our observation regarding the interaction between tRF^{Glu} and YBX1 in vivo, high-throughput sequencing of the YBX1-crosslinked smRNAs (smRNA CLIP-seq) revealed that the majority of these CLIP-seq tags mapped to tRNA loci and represented tRFs (Figures 2A, S2A, and S2B). We observed that YBX1 interacts with a specific subset of tRFs present in these cells. For example, tRF^{GlyTCC} and tRF^{AspGTC}, both with relatively low cellular expression levels, displayed substantial binding to YBX1, whereas highly expressed fragments, tRF^{LysTTT} and tRF^{SerAGA}, were absent among the YBX1 smRNA CLIP-seq tags (Figure S2C). A more global comparison is provided in Figure 2B in which the relative abundance of tRNA fragments mapping to each tRNA locus in MDA-parental cells is shown relative to those from the YBX1 smRNA CLIP-seq. These findings reveal that the interactions between tRFs and YBX1 were not simply a function of tRF abundance in the cell and that YBX1 binding to tRFs is specific and dependent on factors other than tRF levels (e.g., sequence specificity). Based on our YBX1 smRNA CLIP-seq experiment, we identified a number of specific tRFs as most abundantly bound by YBX1; chief among them, the tRNA^{Glu}, tRNA^{Asp}, and tRNA^{Gly} fragments that mapped to the anticodon loops of these tRNAs and a tRNA^{Tyr}-derived fragment matching the intron-containing precursor of this tRNA (Figure 2C).

Transcriptomic profiling under normoxic and hypoxic conditions in both control and YBX1 knockdown cells revealed that in the MDA-parental breast cancer cells, YBX1-bound transcripts were significantly enriched among those downregulated under hypoxia in a YBX1-dependent manner (Figure 2D). This observation suggested the presence of a hypoxia-induced and YBX1-mediated post-transcriptional regulatory program. Gene-expression analysis of non-tumorigenic mammary epithelial MCF10a cells under hypoxia also showed an enrichment of YBX1-bound transcripts among the hypoxia-induced downregulated genes, further strengthening this hypothesis (Z score = 5.7; data not shown). More importantly, highly metastatic MDA-LM2 cells did not exhibit YBX1-dependent downregulated expression of the target transcripts under hypoxia, consistent with their lack of induction of specific hypoxia-induced tRFs (Figure 2D). Given that YBX1 did not show a significant change in expression in MDA-LM2 cells relative to the MDA-parental line (Figure S2D), we hypothesized that the observed enrichment may be mediated by tRFs, which are induced in poorly metastatic but not highly metastatic cells (Figure S1A).

Our findings described above reveal a direct physical interaction between tRNA^{Glu} fragments and YBX1 (Figures 1D and 1E). In order to validate the in vivo interaction between the other tRFs and YBX1, we developed a cell-based competition experiment, based on quantitative PCR (qPCR) assays of smRNAs, in which chemically synthesized tRF mimetics were used to compete with endogenous fragments for YBX1 binding in vivo. We designed specific primers for reliable qPCR-mediated detection of tRF^{Glu}, tRF^{Asp}, and tRF^{Gly} in this competition assay. Under

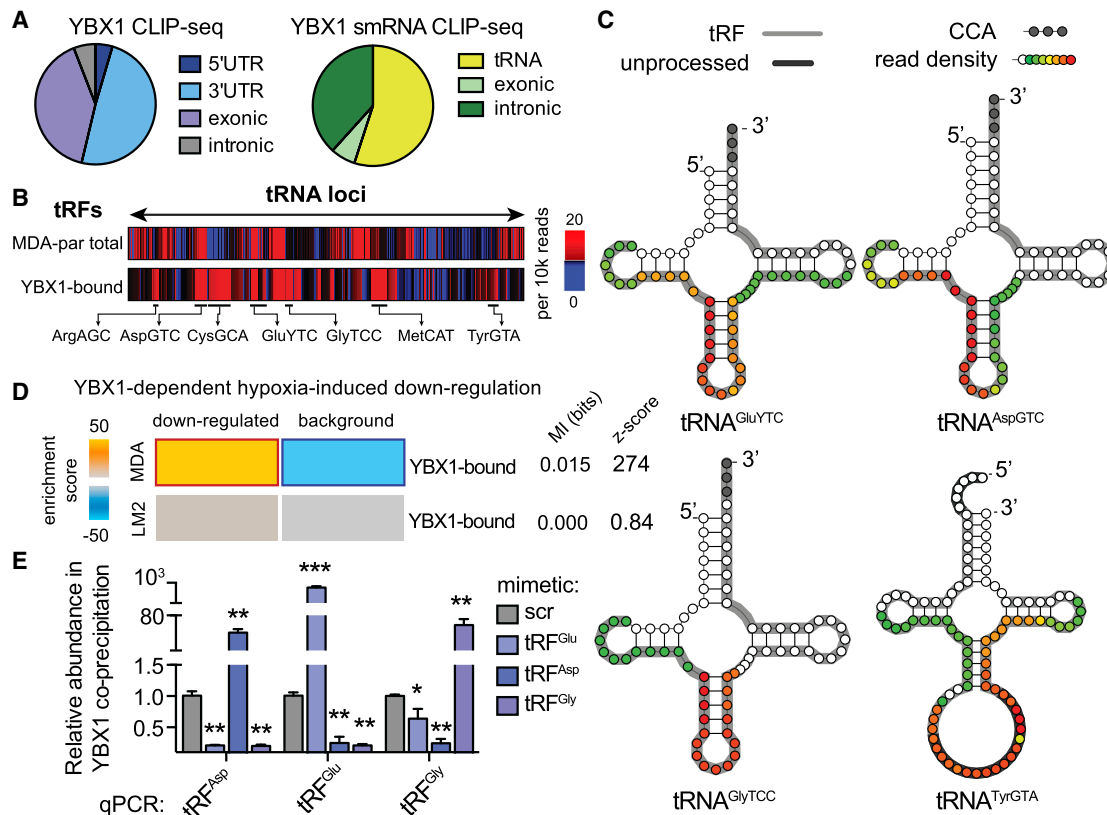


Figure 2. Endogenous YBX1 Interacts with a Large Regulon of Transcripts and smRNAs In Vivo

(A) Pie charts depicting the annotation of YBX1-binding sites obtained from immunoprecipitation of endogenous YBX1 from RNase-treated lysate of UV-irradiated MDA-parental cells followed by high-throughput sequencing of both long- and smRNAs.

(B) Relative frequency of reads mapped to each tRNA locus in MDA-parental smRNA sequencing and YBX1 smRNA CLIP-seq. The tRF species most abundantly bound by YBX1 are marked.

(C) Based on the YBX1 smRNA CLIP-seq results, four species of tRFs bound by YBX1 in vivo were identified. Shown are examples of tRNA structures for each species depicting the boundaries of the identified tRFs along with the YBX1-binding region based on smRNA YBX1 CLIP-seq read density at each position (also see Figure S2A). The gray nucleotides at the 3' end mark the presence of terminal CCA sequences. The dark gray highlights for tRNA^{Tyr}GTA mark the leader and intronic sequences in the unprocessed tRNA. The longest identified forms of each tRF based on our high-throughput sequencing results are also indicated (gray highlight) along with the YBX1 smRNA CLIP-seq read density at each position (overlaid as a heatmap) indicating the YBX1-binding site.

(D) Gene-expression profiling of control and YBX1 knockdown cells was performed under normal and hypoxic conditions in both MDA-parental and MDA-LM2 backgrounds. The set of transcripts that was downregulated under hypoxia in a YBX1-dependent manner was identified for each cell line. Although YBX1-bound transcripts were significantly enriched among YBX1-dependent hypoxia-induced downregulated transcripts in MDA-parental cells, this enrichment was absent in the highly metastatic MDA-LM2 cells.

(E) qPCR-based validation of interactions between YBX1 and tRF^{Asp}GTC, tRF^{Glu}YTC, and tRF^{Gly}TCC. Cells transfected with exogenous tRF mimetics were subjected to UV crosslinking and YBX1 immunoprecipitation. The abundance of each tRF in the co-immunoprecipitated RNA population was then measured using a smRNA qPCR-based quantitation assay (n = 3–4).

Statistical significance is measured using a one-tailed Student's t test: *p < 0.05, **p < 0.01, and ***p < 0.001. Error bars in all panels indicate SEM unless otherwise specified.

the assumption that exogenous tRFs effectively bind YBX1 in vivo (as was shown for tRF^{Glu}), we predicted that the increase in cellular levels of a specific tRF would result in the subsequent displacement of other tRNA fragments from the endogenous YBX1-bound RNA population. To test this hypothesis, we UV-irradiated cells transfected with synthetic tRFs or scrambled controls, immunoprecipitated YBX1 under stringent CLIP-seq conditions (Ule et al., 2005), and used qPCR to detect the abundance of each specific tRF in the YBX1-bound fraction. Consistent with active binding of YBX1 to synthetic tRFs, we observed that exogenous transfection of a given tRF led to depletion of the other as-

sayed endogenous tRF species in every case (Figure 2E). This quantitative assay demonstrates not only that these tRF species bind YBX1 but also that they compete for YBX1 binding in vivo. Moreover, as was the case for tRF^{Glu}, we used 3'-biotinylated short oligonucleotides mimicking tRF^{Asp} and tRF^{Gly} to co-immunoprecipitate interacting proteins from total cell lysates. We observed significant enrichment in endogenous YBX1 protein levels upon co-immunoprecipitation of the tRFs relative to a 3'-biotinylated scrambled oligonucleotide (Figure S2E). Importantly, we also observed a significant upregulation in tRF^{Glu}, tRF^{Asp}, and tRF^{Gly} under hypoxic conditions, quantified using tRF-specific

qPCR in both MDA-parental breast cancer cells and MCF10a non-transformed mammary epithelial cells (Figures S2F and S2G). Consistent with our prior findings, this induction was absent in metastatic MDA-LM2 cells (Figure S2H). While these YBX1-binding tRFs are constitutively expressed, their levels are enhanced in the context of hypoxic stress. This observation indicates that the tRFs identified here can be categorized as tRNA-derived stress-induced RNAs (tiRNAs) and likely play roles in stress responses. More importantly, the absence of their induction in highly metastatic MDA-LM2 cells highlights the potential suppressive roles they may play during breast cancer metastasis, in which tRFs must be antagonized for metastasis to progress.

tRF-Mediated Post-Transcriptional Modulation through YBX1

Previous studies have established a role for another class of tRNA fragments—tRNA 5'-halves (e.g., 5'-tiRNA^{Ala} and 5'-tiRNA^{Cys})—in translation inhibition. These fragments were found to cause translation initiation factors to disengage from mRNAs (Ivanov et al., 2011). This translational inhibition effect was shown to be YBX1 dependent. The molecular mechanism through which this previously identified class of tRFs modulates YBX1 interaction with translation initiation factors remains to be elucidated. YBX1 has also been implicated in other post-transcriptional regulatory programs, most notably transcript stability. We set out to test whether functional interactions between tRFs and YBX1 affect expression levels of endogenous transcripts. We envisioned two plausible molecular mechanisms through which tRFs may affect YBX1 binding to mRNAs in vivo (Figure S3A). First, in a tRF-mediated transcript engagement model, a given tRF may act as a guide RNA whereby the YBX1-tRF complex binds specific target transcripts based on sequence complementarity to the bound tRF, in a manner similar to miRNA-mediated binding of transcripts by Argonaute. In this scenario, reducing tRF levels would lead to reduced YBX1 binding of transcripts. Second, in a tRF-mediated transcript displacement model, YBX1 would interact with tRFs and mRNAs alike, in which case tRFs would be actively competing with endogenous transcripts for YBX1 binding. In this model, reducing tRF levels would lead to greater YBX1 binding of its target mRNAs. Importantly, in the transcript engagement model, mRNAs carrying the reverse-complement of the tRF sequence would be affected in terms of YBX1-binding abundance or expression, whereas in the transcript displacement model, only the transcripts that contain YBX1-binding motifs would be affected by tRF modulation. In order to distinguish between these two molecular mechanisms, we utilized synthetic anti-sense locked-nucleic acids (LNAs) targeting the YBX1-binding site (based on the smRNA CLIP-seq peaks; Figures S2A and S2B) on the most abundantly bound tRFs—namely, tRF^{Asp}, tRF^{Gly}, tRF^{Glu}, and tRF^{Tyr} (Figure 2B). We used specific anti-sense LNAs to bind and inhibit the endogenous forms of these tRFs individually in order to observe their effects on the transcriptome relative to a scrambled LNA control. Importantly, to specifically focus on transcripts impacted by tRF inhibition through their direct interaction with YBX1, we conducted whole-transcriptome profiling of both control and YBX1 knock-down cells transfected with anti-sense LNAs (Figure S3B).

We first sought to identify the YBX1-binding sequences on these four tRFs in order to observe the behavior of mRNAs carrying these sequences (transcript displacement model) or their complementary sequences (transcript engagement model). We made no assumptions about the YBX1-binding site on RNAs, instead opting to test all possible 8-mers (and their reverse complements) along the identified binding sites on each of the four tRFs (sequences shown in Figure 3A). Each 8-mer was assessed for (1) its enrichment (or that of its reverse complement sequence) in 3' UTRs of transcripts that were deregulated in a YBX1-dependent manner in each experiment (see Experimental Procedures) and (2) its enrichment among YBX1-binding sites on endogenous transcripts (YBX1 CLIP-seq). For each tRF, we successfully identified an 8-mer that was enriched in both (1) the 3' UTRs of transcripts upregulated (not downregulated) in a YBX1-dependent manner and (2) the YBX1 CLIP-sites on endogenous mRNAs (Figures 3A and 3B). Given that in each case, it was the specific 8-mer, rather than its reverse complement, that was functionally bound by YBX1 in vivo, the YBX1 modulation of transcript levels observed here is consistent with the model wherein tRFs displace transcripts from YBX1. In this competition-based model, the binding of specific tRFs to YBX1 is inhibited by LNA transfection, allowing free YBX1 to interact with YBX1-binding sites on endogenous transcripts. Increased YBX1 binding would result in higher transcript abundances, most likely through enhanced mRNA stabilization by YBX1 in vivo. Subsequently, we provide additional genomic, molecular, and biochemical evidence that supports this YBX1-dependent post-transcriptional mode of regulation.

In order to determine a consensus binding site for YBX1 on tRFs and mRNAs, we performed multiple alignments for the 8-mers that were independently identified for each tRF. The resulting sequence motif, named CU-box based on the prominence of a C and U at the second and third positions along the identified regular expression representation of the element (Figure 3C), was significantly enriched among CLIP-seq tags in both YBX1 smRNA ($p < 0.002$) and long RNA (mRNA) CLIP-seq datasets ($p < 10^{-80}$; Figure 3D). Importantly, this element resembles an in vitro YBX1-binding motif described previously (Figure S3C). Moreover, tracking cross-linking-induced mutation sites (CIMS; Zhang and Darnell, 2011), which mark protein-RNA interactions at single-nucleotide resolution, at and around CU-box elements (10 flanking nucleotides) across the YBX1 CLIP sites revealed that the cross-linked nucleotides were most frequent at the site of the motif (Figure S3D). This observation further supports a direct physical interaction between YBX1 and CU-box elements. Taken together, our results indicate that YBX1 interacts with both endogenous target transcripts and tRFs via the CU-box element. It should also be noted that the SCUBYC motif identified in Figure 1A constitutes a specific subset of the CU-box element described here (CompareACE score of 0.85). The commonality of the binding site would enable tRFs to competitively modulate the levels of YBX1 available for transcript binding, which would in turn affect the expression of a large set of target transcripts.

Competitive tRF Binding to YBX1 Results in Transcript Destabilization

In addition to the anti-sense LNA-mediated tRF loss-of-function experiments followed by transcriptomic profiling, we also

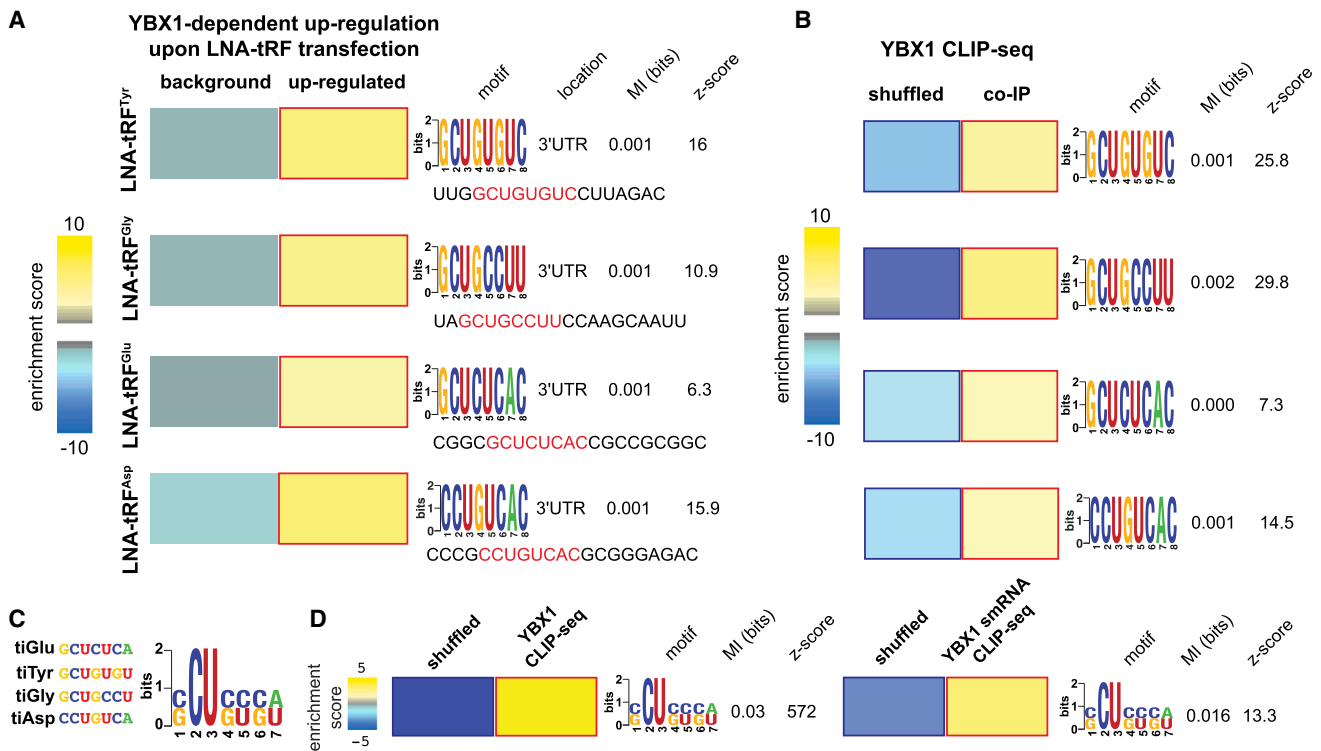


Figure 3. YBX1 Interacts with Long- and smRNAs via a Specific Linear Sequence Motif

(A) The transcripts upregulated upon anti-sense LNA transfections targeting each of the four identified tRFs were compared to the remainder of the transcriptome (background) to identify over-representation of specific sequence elements in their 3' UTRs. Here, we have shown the enrichment of specific 8-mers along each YBX1-bound tRF in the 3' UTRs of these transcripts as a heatmap, with yellow and blue showing the extent of enrichment and depletion, respectively (red and blue borders mark statistical significance). Also shown are the associated mutual information values and Z scores (Elemento et al., 2007). We have provided the sequence of each tRF and highlighted the identified 8-mers.

(B) These 8-mers were also required to be enriched among the YBX1-binding sites identified with CLIP-seq. We used a shuffled version of each YBX1-binding site to create a background set and tested the enrichment of each 8-mer in the YBX1-binding sites relative to shuffled controls.

(C) In order to infer a consensus element for YBX1 on these tRFs, the four significant 8-mers were aligned, and the possible nucleotides at each position were combined to build the CU box element represented as a regular expression.

(D) The CU box motif showed a significant enrichment in both long- and smRNA YBX1 CLIP-seq data sets relative to randomly shuffled sequences, indicating that YBX1 binds a common linear sequence motif on both short and long endogenous RNAs.

Error bars in all panels indicate SEM unless otherwise specified.

transfected control and YBX1 knockdown cells with synthetic tRF mimetics in gain-of-function experiments delineated in Figure S3E. For this, we performed two separate experiments; in one, we transfected control and YBX1 knockdown cells with the mimetics representing the long form of each identified tRF (Figures 2C and S2B), and in another, we used ~20 nucleotide short synthetic tRF mimetics containing the identified YBX1-binding sites. Although YBX1-bound transcripts (and transcripts with CU boxes in their 3' UTRs) were significantly upregulated upon LNA transfections (Figures 4A, S4A, and S4B), these transcripts were significantly downregulated in the context of tRF mimetic transfection in a YBX1-dependent manner (Figures 4B, 4C, S4C, and S4D). These observations reveal that (1) exogenously transfected tRF mimetics act as modulators of the YBX1 regulon, and (2) short tRF mimetics carrying the YBX1-binding site are sufficient for exerting this regulatory effect.

To determine whether the observed YBX1-dependent tRF-mediated modulations in YBX1 target transcripts' levels were occurring post-transcriptionally, we performed whole-genome

transcript stability measurements. Through α -amanitin-based inhibition of RNA polymerase II in anti-sense LNA-transfected control and YBX1 cells, we found that the observed increase in YBX1-targeted transcript abundance resulted from a significant enhancement of their stabilities in a YBX1-dependent manner (Figures 4D, S4E, and S4F). This observation further establishes the role of these YBX1-tRF interactions as a coherent and functional post-transcriptional regulatory program (Figure S4G). Importantly, the changes observed in the expression of YBX1-dependent transcripts upon transfection of anti-sense LNAs or tRF mimetics were significantly anti-correlated (Figure S4H). Further supporting this model, we also observed a reduction in total RNA bound to YBX1 upon transient transfection of an exogenous tRF^{Glu} mimetic (Figure S5A).

Target-Specific Validation of a Role for tRF-YBX1 in Transcript Destabilization

Whole-genome expression and stability measurements support a transcript displacement model for tRF-YBX1 interaction,

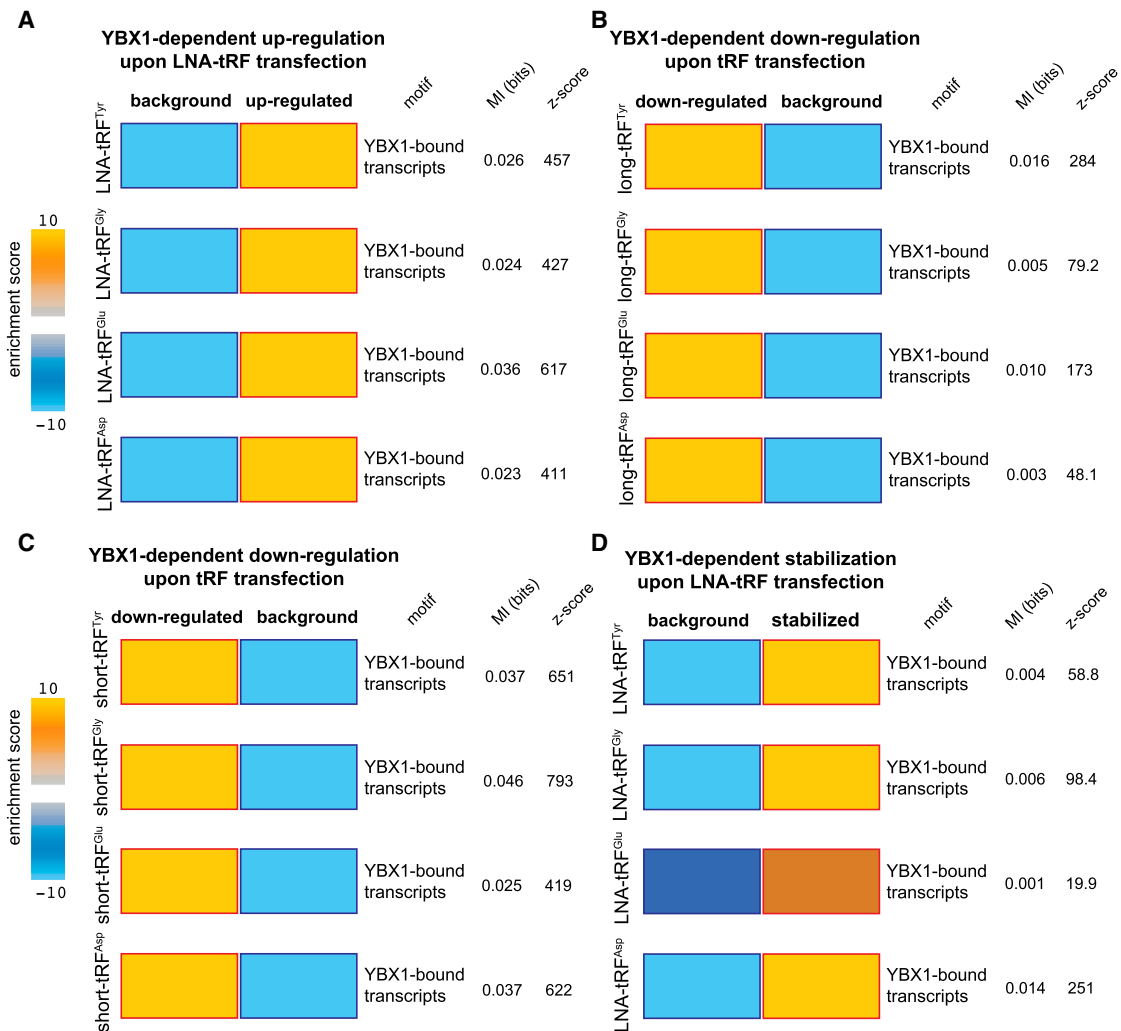


Figure 4. Endogenous Transcripts Bound by YBX1 Are Modulated by YBX1-Bound tRFs

In order to measure the post-transcriptional regulatory consequences of tRFs, gain-of-function and loss-of-function experiments were performed by transfecting synthetic tRF mimetics or inhibitory anti-sense LNAs, for each of the four YBX1-binding tRFs, in normal and YBX1 knockdown cells. Transcripts that were up- or downregulated in a YBX1-dependent manner were identified by comparing the gene-expression changes in normal cells relative to those in YBX1 knockdown cells (Figure S3). Transcripts that interact with YBX1 in vivo (determined from YBX1 CLIP-seq data) were significantly de-regulated upon modulations of tRF levels: (A) they were upregulated upon LNA-mediated inhibition of YBX1-binding tRFs; (B and C) they were downregulated in the presence of exogenously added short and long tRF mimetics (~60 and ~20 nucleotides, respectively; see Figure S2B), and (D) the observed upregulation in the LNA-transfected cells coincided with a significant increase in their stability. Whole-genome transcript stability measurements were performed in LNA-transfected cells using α -amanitin-mediated inhibition of RNA polymerase II followed by RNA extraction and profiling at 0 and 8 hr time points. In all data sets, the calculated mutual information values (in bits) and their associated p values are provided. Also shown are the enrichment scores, presented as $\log P$ (positive for enrichments and negative for depletions), where P is calculated from hypergeometric distribution (shown as a heatmap with blue and gold showing depletion and enrichment, respectively). The red and blue borders mark statistical significance of the enrichment/depletions. Error bars in all panels indicate SEM unless otherwise specified.

wherein tRFs effectively compete with endogenous transcripts for YBX1 binding. In order to independently validate our observations for a specific set of targets, we chose HMGA1, CD151, CD97, and TIMP3 for target-specific follow-up experimental validation based on pervasive in vivo interactions with YBX1 along their 3' UTRs (Figure 5A). Transfection of tRF-specific anti-sense LNAs significantly upregulated and stabilized these transcripts (Figures 5B and 5C). More importantly, their observed upregulation was abrogated in YBX1 knockdown cells (Figures 5B and 5C).

As mentioned earlier, under hypoxic conditions, in which the YBX1-bound tRFs were shown to be upregulated in MDA-parental cells (Figure 2C), we observed a concomitant downregulation of YBX1 target transcripts (Figure 2D). Importantly, this hypoxia-induced YBX1-dependent downregulation, which was absent in highly metastatic MDA-LM2 cells, was diminished once hypoxic cells were transfected with anti-sense LNAs (Figure S5B). We also observed a similar expression pattern for HMGA1, CD97, and TIMP3 transcripts when tested under hypoxia and normoxia in both control and YBX1 knockdown cells

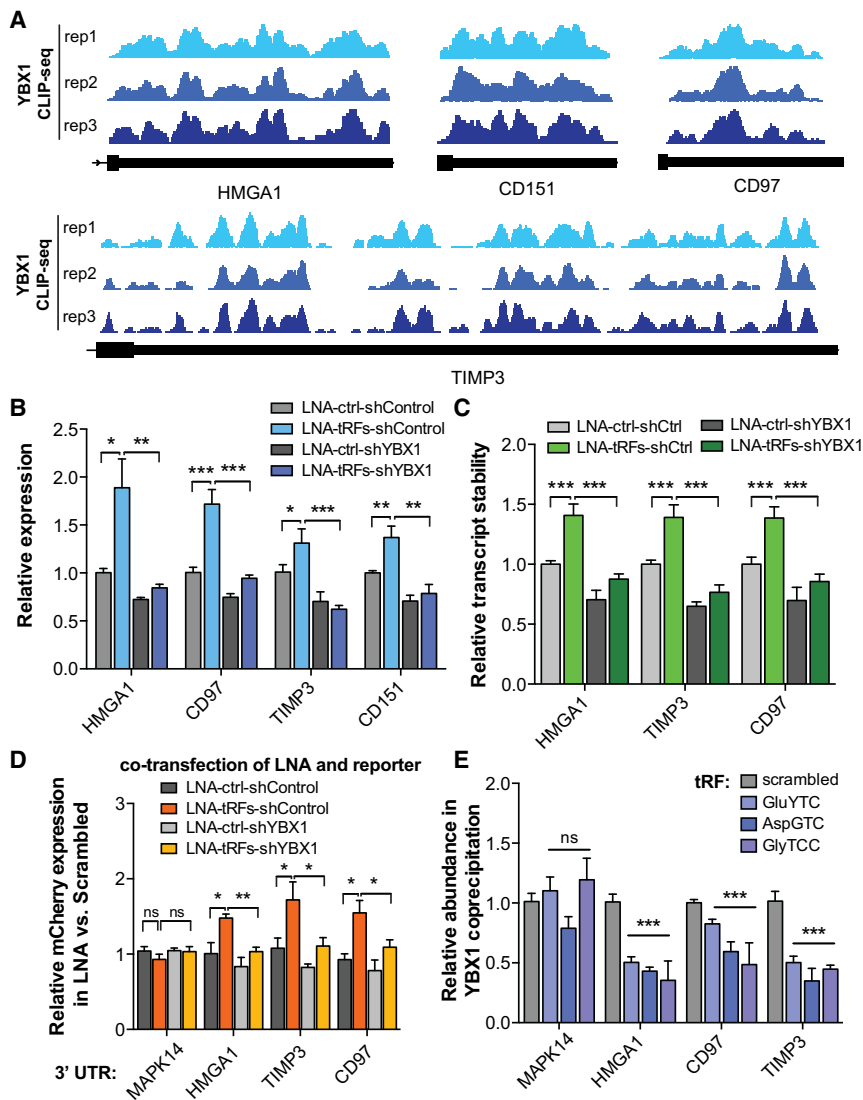


Figure 5. YBX1 Target Transcripts and Their Response to Changes in tRF Levels

(A) YBX1 interacts with the 3' UTRs of HMGA1, CD151, CD97, and TIMP3. The last exon of the indicated transcripts are shown with mapped reads from experimental replicates of YBX1 CLIP-seq.

(B) Transfection of anti-sense LNAs against YBX1-binding tRFs resulted in the upregulation of HMGA1, CD151, CD97, and TIMP3 transcripts, in a YBX1-dependent manner, as determined by qPCR measurements.

(C) Similarly, transfecting anti-sense LNAs resulted in a significant stabilization of HMGA1, CD97, and TIMP3 transcripts in a YBX1-dependent manner. Whole-genome RNA stability measurements were performed using α -amanitin-mediated inhibition of RNA polymerase II (see [Experimental Procedures](#)).

(D) A GFP/mCherry dual-reporter assay was used to measure the effects of cloning HMGA1, CD97, and part of the TIMP3 3' UTRs downstream of mCherry using qRT-PCR. The 3' UTR of MAPK14, which is devoid of YBX1 tags, was included as a control. Consistent with our prior findings, LNA transfections resulted in a significant increase in relative mCherry expression in a YBX1-dependent manner.

(E) Exogenously added tRF mimetics, although showing no effect on MAPK14 abundance, resulted in a significant depletion of HMGA1, CD97, and TIMP3 transcripts from the YBX1 co-immunoprecipitated RNA population.

Statistical significance is measured using a one-tailed Student's *t* test: **p* < 0.05, ***p* < 0.01, and ****p* < 0.001. Error bars in all panels indicate SEM unless otherwise specified.

tRF-Mediated Modulation of Cancer Progression and Metastasis

YBX1 has been implicated in cancer progression. YBX1 overexpression is correlated with tumorigenic phenotypes and has also been shown to promote cancer

in MDA-parental and MDA-LM2 backgrounds (Figure S5C). Consistently, we found that the 3' UTR sequences of these transcripts cloned downstream of a bi-directional reporter construct were sufficient to confer YBX1-dependent upregulation of the reporter upon transfection of anti-sense LNAs targeting the four selected tRFs (Figure 5D). The MAPK14 3' UTR, which was not bound by YBX1 as assessed by CLIP-seq, served as a comparative control. Importantly, the YBX1 dependence of this tRF-mediated response to LNA transfection further supports a role for YBX1 in stabilizing target transcripts through binding of 3' UTR elements (Figure 5D).

Consistent with our proposed tRF-mediated transcript displacement model, we also observed that transfecting tRF mimetics into cells followed by YBX1 co-immunoprecipitation depleted HMGA1, TIMP3, and CD97 transcripts from the YBX1-bound RNA population (Figure 5E). This observation further supports direct competition between tRFs and endogenous transcripts for YBX1 binding in vivo.

metastasis (Jürchott et al., 2010; Matsumoto and Bay, 2005; Uchiyama et al., 2006; Wu et al., 2012). Given the role of tRFs in suppressing the expression of YBX1 target genes, we hypothesized that this class of smRNAs may act to suppress cancer progression. For example, a tRF-YBX1 signature based on the average expression of roughly 70 YBX1-bound transcripts with robust modulations in response to anti-sense LNA or tRF mimetic transfections was found to be significantly associated with cancer stage as well as relapse-free survival of patients with breast cancer (Figures S6A and S6B). The entire YBX1-tRF regulon—defined as the subset of endogenous transcripts that are bound by YBX1 in vivo and whose expression is modulated by the transfection of tRF mimetics and anti-sense LNAs in a YBX1-dependent manner—contains hundreds of transcripts, including many known promoters of tumorigenesis and metastasis, such as AKT, EIF4G1, ITGB4, and HMGA1 (Figures 6A and S6C). We also noted highly significant associations between increased expression of multiple tRF-YBX1 targets that are key

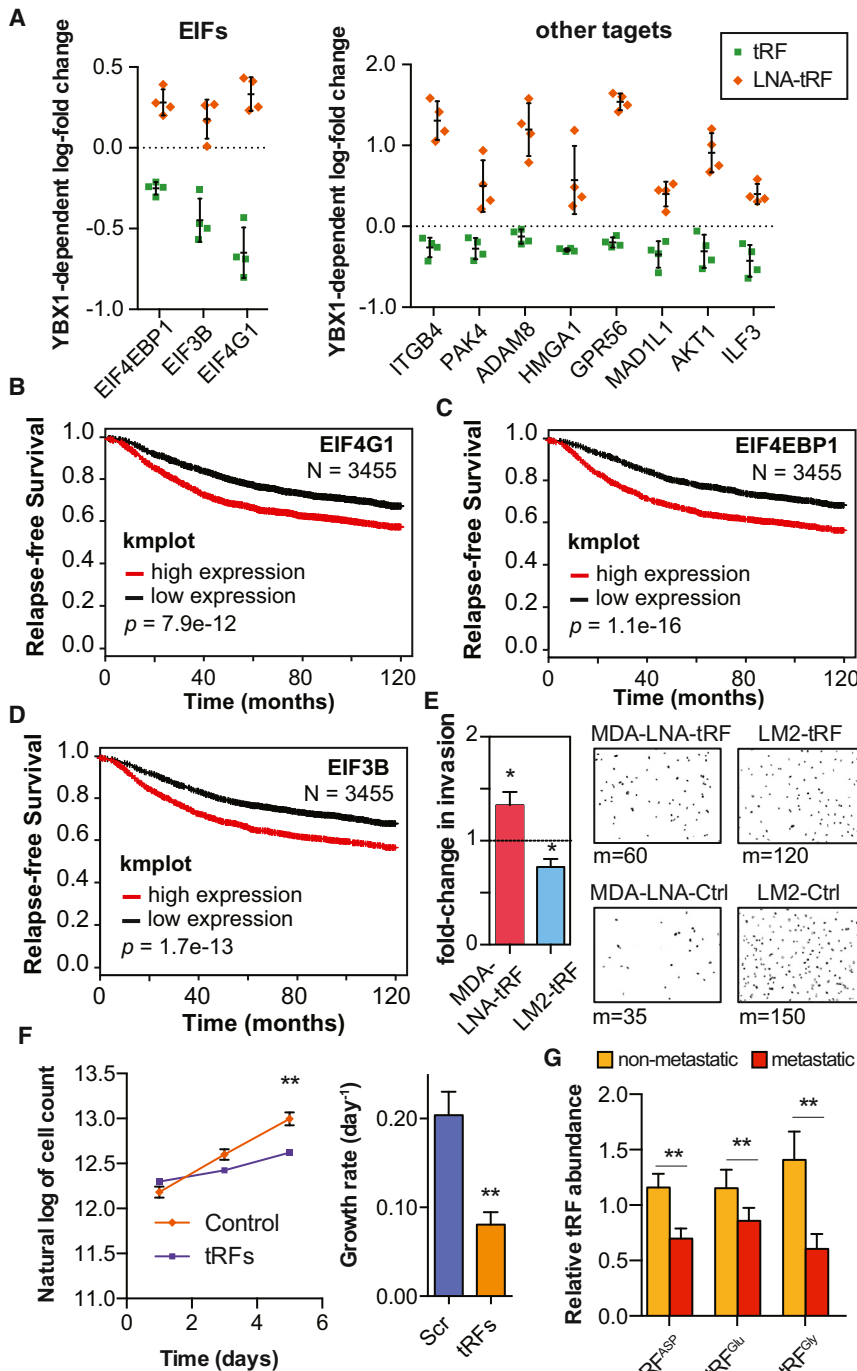


Figure 6. YBX1-Binding tRFs Play a Significant Role in Modulating Oncogenes

(A) Competitive displacement of YBX1 from its target transcripts by tRFs resulted in the down-regulation of a large set of oncogenes and metastasis promoter genes.

(B–D) Kaplan-Meier curves for three translation initiation factors that were modulated by tRFs via YBX1 binding (Gyorffy et al., 2012).

(E) Exogenously added tRF mimetics or anti-sense LNAs resulted in a significant increase and decrease in cancer cell invasion, respectively. Shown are the fold-changes in cancer cell invasion for MDA-parental cells transfected with LNAs and MDA-LM2 cells transfected with tRFs. We have also included representative fields from the invasion inserts along with the median of cells observed in each cohort (n = 7–8).

(F) Growth rates (estimated based on an exponential model) under serum-starved conditions for MDA-LM2 cells transfected with tRF mimetics relative to mock-transfected cells (n = 6).

(G) qRT-PCR assays were used to quantify the levels of tRF^{ASP}, tRF^{GLU}, and tRF^{GLY} in metastatic (n = 18) and non-metastatic (n = 9) primary breast cancers.

For comparing growth under serum-starved conditions, two-way ANOVA was used to measure statistical significance. For all other cases, statistical significance was measured using one-tailed Student's t test: *p < 0.05, **p < 0.01, and ***p < 0.001. Error bars in all panels indicate SEM unless otherwise specified.

sion capacity in vitro (Figures 6E and S6D). Conversely, transfection of tRF mimetics into metastatic MDA-LM2 and CN-LM1a lines significantly reduced cancer cell invasion (Figures 6E and S6D). We also observed a substantial decrease in cell proliferation rate under serum-starved conditions in the presence of these tRFs, further highlighting their roles as components of a general stress-response pathway (Figure 6F). These tRFs were ineffective at suppressing these phenotypes in cells depleted of YBX1—consistent with a required role for YBX1 in modulating these tRF-mediated responses (Figures S6E and S6F).

Importantly, we also detected tRF^{ASP}, tRF^{GLU}, and tRF^{GLY} in RNA samples from

metastatic and non-metastatic primary tumors as well as normal breast tissue. Consistent with a tumor-suppressive role for these tRFs, their levels were significantly lower in breast cancer tissue relative to normal breast tissue (Figure S6G). Moreover, if these specific tRFs suppress metastatic progression, we would expect that there would be a selection for reduced expression of these tRFs during this process. Indeed, we observed a significant trend toward reduced tRF levels in

components of translation initiation (EIF4G1, EIF4EBP1, and EIF3B) and reduced relapse-free survival ($p = 8 \times 10^{-12}$, 1×10^{-16} , and 2×10^{-13} , respectively; n = 3455; Figures 6B–6D). The transcripts of these oncogenes, which play roles in various aspects of cellular function, including translation and cell signaling, were repressed by tRFs in breast cancer cells (Figure 6A). Transfection of anti-sense LNAs targeting the tRFs into MDA-parental and CN-parental cancer cells significantly enhanced cell inva-

metastatic samples compared to primary tumors ($p = 0.003$, $n = 27$; [Figure 6G](#)).

To demonstrate the physiological relevance of this hypoxia-induced tRF-YBX1 pathway, we used a reporter driving the expression of luciferase under a hypoxia response promoter ([Figure S7A](#)). Consistent with MDA-parental breast cancer cells experiencing hypoxia early in the metastatic process in the lungs of xenografted mice, cells carrying this reporter exhibited induction of the hypoxia-induced pathway 24 hr post-injection ([Figure S7B](#)). To probe the *in vivo* expression of tRF-YBX1 targets, we constructed a lentiviral system with firefly luciferase reporter fused to 3' UTRs of CD97 and TIMP3 (tRF-YBX1 targets) as well as MAPK14 (as a control). Consistent with YBX1 stabilizing these transcripts by binding to their 3' UTRs, immediately after injection, we observed a significantly lower luciferase activity for CD97 and TIMP3 3' UTRs in YBX1 knockdown cells ([Figure S7C](#)). More importantly, the CD97 and TIMP3 reporters showed significantly lower day 3 to day 0 luciferase activity ratios in control cells compared to YBX1 knockdown cells ([Figure S7D](#)). These findings are consistent with the *in vivo* induction of tRFs under hypoxia and the YBX1 dependence of the associated response. It should be highlighted that this reduction was absent in the luciferase reporter carrying the control MAPK14 3' UTR ([Figure S7D](#)).

Consistent with the observed clinical associations and our *in vitro* as well as *in vivo* findings, transfection of tRF mimetics and anti-sense LNAs significantly impacted metastatic colonization in *in vivo* lung colonization assays by multiple independent cell lines. MDA-parental cells transfected with anti-sense LNAs exhibited significantly higher metastatic colonization activity relative to cells transfected with scrambled controls ([Figures 7A and S7E](#)). Similarly, transfection of inhibitory anti-sense LNAs in CN-parental cells also caused a marked and significant increase in metastatic colonization of the lungs ([Figures 7B and S7E](#)). In contrast, exogenous transfection of tRF mimetics into highly metastatic lines (MDA-LM2 and CN-LM1a cells) significantly reduced cancer metastasis to the lungs ([Figures 7C, 7D, and S7E](#)). We also tested and validated the impact of tRF modulation on the metastatic activity of a third human breast cancer cell-line—HCC1806 ([Figures 7E and S7E](#)). It should be noted that in these *in vivo* lung colonization assays, a clear and significant difference in the normalized signal could be detected early in the *in vivo* experiments ([Figure S7F](#)). This early impact on metastasis is in part consistent with the role of tRFs in cancer cell invasion, which is an early determinant of metastatic progression. This early difference persists throughout the experiment despite the dilution of mimetics and anti-sense LNAs ([Figure S7G](#)), and the two cohorts fail to converge. Importantly, consistent with a YBX1-dependent mode of action, depleting YBX1 from cancer cells using RNAi made cells insensitive to tRF-mediated modulation of metastatic activity ([Figure S7H](#)).

DISCUSSION

By integrating biochemical, molecular, computational, and phenotypic analyses, we have found that a specific set of tRFs functionally engages the oncogenic RNA-binding protein YBX1. These fragments, which contain a CU box motif, post-transcriptionally suppress the expression of YBX1 transcripts

by competitively displacing them from YBX1. We find that YBX1 binds and promotes the stability of a large set of transcripts, thus modulating a large regulon with broad consequences for cellular function. Our study reveals the first comprehensive and *in vivo* interaction map between YBX1, one of the most overexpressed oncogenes observed in human cancer (upregulated in 10% of all cancer versus normal tissue data sets; Oncomine), and its post-transcriptional target transcripts ([Lasham et al., 2012](#); [Uchiumi et al., 2006](#); [Wu et al., 2012](#)). A number of these transcripts encode established drivers of oncogenesis, such as EIF4G1, ITGB4, AKT1, and ADAM8. YBX1 stabilization of oncogenic transcripts is mediated by its binding to CU box motifs, which are primarily located in the 3' UTRs of transcripts. The displacement of oncogenic transcripts from YBX1 results in their destabilization and downregulation. Consistent with a tumor-suppressive role for these YBX1-antagonistic smRNAs, their introduction into breast cancer cells inhibited breast cancer growth under serum starvation, cell invasion, and metastasis. Conversely, inhibiting these fragments by anti-sense LNAs enhanced these phenotypes.

This tRF-mediated displacement mechanism of post-transcriptional silencing involving the binding of specific tRFs to YBX1 differs from RNAi-mediated silencing in that small tRNA fragments do not serve as guides for transcript engagement by YBX1. Rather, they competitively displace and thus destabilize YBX1-bound transcripts. Our findings expand the repertoire of endogenous smRNA-mediated post-transcriptional modes of regulation that have been previously described (RNAi, microRNA, and CeRNA; [Karaca et al., 2014](#); [Lujambio and Lowe, 2012](#); [Salmena et al., 2011](#)).

Our findings reveal that a specific set of fragments contain tumor-suppressive and metastasis-suppressive activity. We propose that these fragments are generated as a result of oncogenic stress as an internal mechanism for tumor suppression. We speculate that during breast cancer evolution, two mechanisms counter this smRNA-mediated tumor-suppressive mechanism: the first being the evasion of the hypoxia-evoked induction of tumor-suppressive tRFs, and the second being YBX1 upregulation. Consistent with these findings, tRNA fragments were detected at significantly lower levels in metastatic breast cancer relative to non-metastatic cancers, whereas YBX1 is known to be upregulated as a function of cancer progression ([Lasham et al., 2012](#)).

We find that the repressive effects of these endogenous tRNA fragments on the abundance and stability of oncogenic transcripts is moderate in scale. Nonetheless, our loss-of-function and gain-of-function studies involving these fragments reveal robust *in vitro* and *in vivo* effects resulting from their modulation in breast cancer. We believe that these effects result from the coordinated post-transcriptional control of a large set of YBX1-dependent oncogenes whose concomitant suppression results in robust phenotypic effects. These observations parallel those seen with microRNAs implicated in cancer progression—moderate post-transcriptional suppressive effects on groups of transcripts involved in common oncogenic processes ([Lujambio and Lowe, 2012](#)).

An association between YBX1 and tRNA halves has been previously described ([Emara et al., 2010](#)). Transfection of tRNA

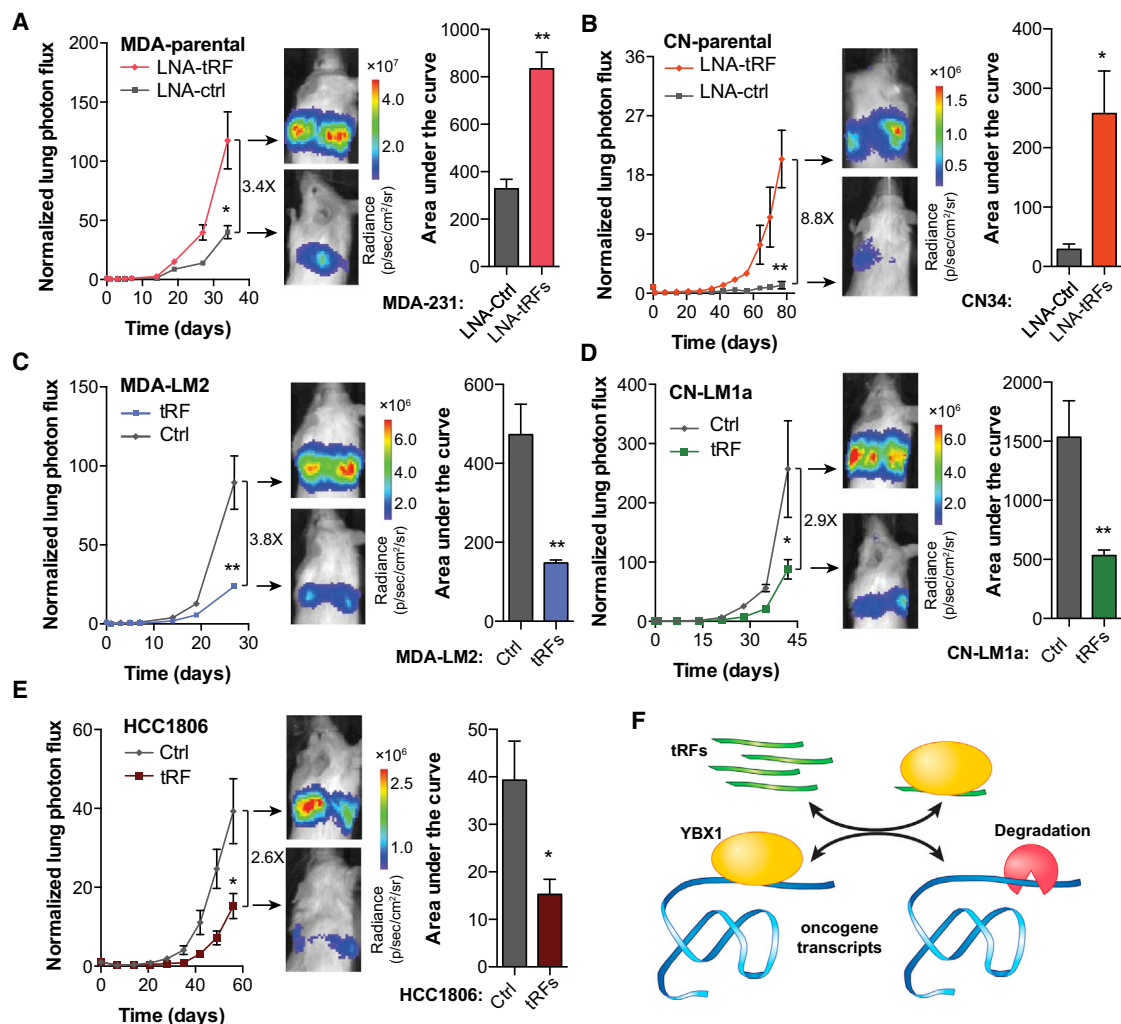


Figure 7. YBX1-Binding tRFs Play a Significant Role as Suppressors of Tumor Progression and Metastasis

(A and B) Bioluminescence imaging plot of metastatic lung colonization by MDA-parental and CN-parental cells transfected with synthetic anti-sense LNAs against all four YBX1-binding tRFs. Representative images along with quantification of the area-under-the-curve for each mouse are also included ($n = 3-5$ in each cohort).

(C and D) Bioluminescence imaging plot of lung metastasis by MDA-LM2 and CN-LM1a cells transfected with the four YBX1-binding tRFs. Representative images and area-under-the-curve quantifications are also included ($n = 4-5$ in each cohort).

(E) Bioluminescence imaging plot of lung metastasis by HCC1806 cells transfected with the four YBX1-binding tRFs. Representative images and area-under-the-curve quantifications are also included ($n = 5$ in each cohort).

(F) Schematic of tRF-mediated modulation of invasion and metastatic lung colonization through in vivo titration of YBX1 and the subsequent destabilization of its oncogenic and pro-metastatic targets.

For comparing metastasis colonization assays, two-way ANOVA was used to measure statistical significance. For all other cases, statistical significance is measured using one-tailed Student's *t* test: * $p < 0.05$, ** $p < 0.01$, and *** $p < 0.001$. Error bars in all panels indicate SEM unless otherwise specified.

halves arising from tRNA^{Ala} and tRNA^{Cys} was found to globally inhibit translation. These tRNA halves were found to repress translation by ~30%, and these effects seemed to result from the disengagement of translational initiation factor EIF4G1. Additionally, it was found that the effect of these tRNA halves on translational repression was YBX1 dependent. These tRNA halves were found to interact with YBX1, and both fragments contained a terminal oligoguanine motif. Although molecular mechanisms for the global inhibition of translation by these tRNA halves and their in vivo roles have yet to be delineated,

the authors proposed that these fragments suppress translation by interfering with EIF4G protein in a YBX1-dependent manner. Our findings described here reveal a distinct mechanism of action by a different class of tRFs. The tRFs we have implicated belong to a distinct class of fragments and mediate post-transcriptional destabilization of a specific set of transcripts by directly engaging YBX1 in a sequence-specific manner. Importantly, our observations regarding the downregulation of elongation initiation factors at the transcript level are in agreement with the broad inhibition of translation reported to be induced by a

broader class of tRNA fragments. Moreover, the different mechanisms by which distinct classes of tRFs regulate YBX1-dependent gene expression highlight the significance of this stress-activated regulator in mammalian gene regulation.

Taken together, our findings support a role for endogenous tRFs in destabilizing oncogenic transcripts through their direct binding to YBX1. tRF binding of YBX1 leads to displacement of endogenous oncogenic transcripts from YBX1—resulting in their destabilization (Figure 7F). Based on this model, specific tRFs mediate a unique post-transcriptional gene-expression regulatory program through their engagement of YBX1. It should be noted, however, that the regulatory interactions mediated by tRFs are unlikely to be limited to YBX1, and they likely serve as a component of a larger regulatory network consisting of various RNA-binding proteins and small ncRNAs. We should also point out the possible role of RNA modifications in the functionality of tRNA fragments. Given that extensive base modifications in tRNAs are crucial for their function, future studies should address the potential role of these modifications in tRNA fragments as well. Two lines of evidence in our data suggest a substantial role for these modifications in modulating the regulatory effects of endogenous tRFs. First, although the induction of endogenous tRFs under hypoxia in MDA-parental cells was substantially more modest than exogenous transfection of synthetic tRF mimetics, the ensuing YBX1-dependent downregulation of the tRF-YBX1 regulon was higher in magnitude for endogenous fragments relative to unmodified transfected mimetics. Second, consistent with the possible importance of RNA modifications in tRFs, transfection of anti-sense LNAs that inhibit endogenous fragments was more potent than that of synthetic mimetics in eliciting a regulatory response in the cells. This higher potency of endogenous tRFs relative to synthetic mimetics could be explained by the presence of RNA modifications that are likely to affect the structure, stability, and binding affinity of these fragments.

Although we have shown that the fragments described here modulate specific phenotypes through transcript displacement from YBX1, they may also modulate the activity of additional *trans* factors. From a broader perspective, we speculate that fragments arising from other classes of ncRNAs, such as ribosomal and sno-RNAs, might mediate similar effects by displacing distinct RNA-binding proteins (or other ncRNAs) from their endogenous downstream targets.

EXPERIMENTAL PROCEDURES

Tissue Culture

HEK293T, MDA-MB-231, and CN34 cells and their derived sub-lines, CN-LM1a and MDA-LM2, were cultured in DMEM-based media supplemented with 10% FBS, glutamine, pyruvate, penicillin, streptomycin, and fungizone. RNAi and DNA transfections were performed using Lipofectamine 2000 (Invitrogen) and TransIT-293 (Mirus), respectively.

Exogenous tRF and Anti-Sense LNA Transfection

tRF anti-sense LNA oligonucleotides (Exiqon) or synthetic tRF mimetics (IDT) were transfected using Lipofectamine 2000 in Reduced Serum Media (Life Technologies) for a final concentration of 50 nM consisting of equal parts of each tRF decoy or anti-tRF LNA. After 6 hr of incubation, transfection media were replaced with fresh media. Cells were subjected to *in vitro* and *in vivo*

studies 48 hr after transfection. The sequences for the short and long tRF mimetics are provided in Figure S2B.

Animal Studies

All mouse studies were conducted according to a protocol approved by the Institutional Animal Care and Use Committee (IACUC) at the Rockefeller University.

ACCESSION NUMBERS

The data for high-throughput sequencing and microarray profiling experiments are deposited at GEO under the accession number GSE63605.

SUPPLEMENTAL INFORMATION

Supplemental Information includes seven figures and Extended Experimental Procedures and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.02.053>.

AUTHOR CONTRIBUTIONS

S.F.T. conceived the project and supervised all research. S.F.T., H.G., X.L., H.C.B.N., L.F., and S.Z. wrote the manuscript. H.G., X.L., H.C.B.N., S.Z., and L.F. designed, performed, and analyzed the experiments.

ACKNOWLEDGMENTS

We are grateful to Saeed Tavazoie, Claudio Alarcon, and Alexander Nguyen for their thoughtful comments on this manuscript. We thank Nora Pencheva for insightful discussions on YBX1. We thank C. Zhao, W. Zhang, C. Lai, and S. Dewell of the Rockefeller Genomics Resource Center for assistance with next-generation RNA sequencing and microarray profiling. We also thank H. Molina at Rockefeller Proteomics Resource Center. H.G. was previously supported by an Anderson Cancer Center Fellowship and is currently the recipient of a Ruth L. Kirschstein National Research Service Award from the NIH (T32CA009673-36A1). S.F.T. is a Department of Defense Era of Hope Scholar and a Department of Defense Breast Cancer Collaborative Scholars and Innovators Award recipient.

Received: September 25, 2014

Revised: January 9, 2015

Accepted: February 27, 2015

Published: May 7, 2015

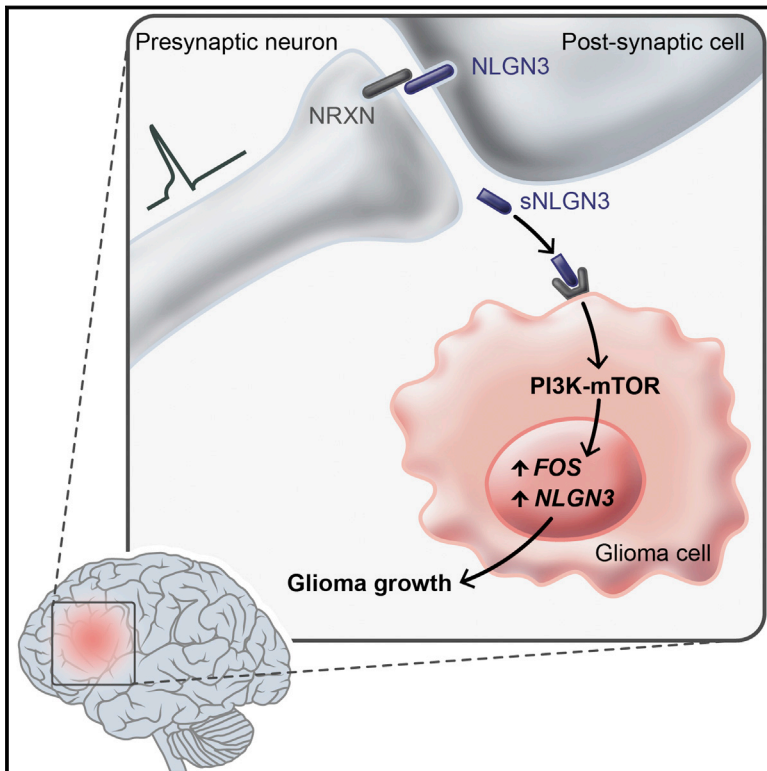
REFERENCES

- Borek, E., Baliga, B.S., Gehrke, C.W., Kuo, C.W., Belman, S., Troll, W., and Waalkes, T.P. (1977). High turnover rate of transfer RNA in tumor tissue. *Cancer Res.* 37, 3362–3366.
- Bristow, R.G., and Hill, R.P. (2008). Hypoxia and metabolism. Hypoxia, DNA repair and genetic instability. *Nat. Rev. Cancer* 8, 180–192.
- Cole, C., Sobala, A., Lu, C., Thatcher, S.R., Bowman, A., Brown, J.W.S., Green, P.J., Barton, G.J., and Hutvagner, G. (2009). Filtering of deep sequencing data reveals the existence of abundant Dicer-dependent small RNAs derived from tRNAs. *RNA* 15, 2147–2160.
- Elemento, O., Slonim, N., and Tavazoie, S. (2007). A universal framework for regulatory element discovery across all genomes and data types. *Mol. Cell* 28, 337–350.
- Emara, M.M., Ivanov, P., Hickman, T., Dawra, N., Tisdale, S., Kedersha, N., Hu, G.F., and Anderson, P. (2010). Angiogenin-induced tRNA-derived stress-induced RNAs promote stress-induced stress granule assembly. *J. Biol. Chem.* 285, 10959–10968.
- Fu, H., Feng, J., Liu, Q., Sun, F., Tie, Y., Zhu, J., Xing, R., Sun, Z., and Zheng, X. (2009). Stress induces tRNA cleavage by angiogenin in mammalian cells. *FEBS Lett.* 583, 437–442.

- Gebetsberger, J., and Polacek, N. (2013). Slicing tRNAs to boost functional ncRNA diversity. *RNA Biol.* *10*, 1798–1806.
- Gyorffy, B., Lanczy, A., and Szallasi, Z. (2012). Implementing an online tool for genome-wide validation of survival-associated biomarkers in ovarian-cancer using microarray data from 1287 patients. *Endocr. Relat. Cancer* *19*, 197–208.
- Ivanov, P., Emara, M.M., Villen, J., Gygi, S.P., and Anderson, P. (2011). Angiogenin-induced tRNA fragments inhibit translation initiation. *Mol. Cell* *43*, 613–623.
- Jurchott, K., Kuban, R.J., Krech, T., Bluthgen, N., Stein, U., Walther, W., Friese, C., Kietbasa, S.M., Ungethum, U., Lund, P., et al. (2010). Identification of Y-box binding protein 1 as a core regulator of MEK/ERK pathway-dependent gene signatures in colorectal cancer cells. *PLoS Genet.* *6*, e1001231.
- Karaca, E., Weitzer, S., Pehlivan, D., Shiraishi, H., Gogakos, T., Hanada, T., Jhangiani, S.N., Wiszniewski, W., Withers, M., Campbell, I.M., et al.; Baylor Hopkins Center for Mendelian Genomics (2014). Human CLP1 mutations alter tRNA biogenesis, affecting both peripheral and central nervous system function. *Cell* *157*, 636–650.
- Krol, J., Loedige, I., and Filipowicz, W. (2010). The widespread regulation of microRNA biogenesis, function and decay. *Nat. Rev. Genet.* *11*, 597–610.
- Lasham, A., Samuel, W., Cao, H., Patel, R., Mehta, R., Stern, J.L., Reid, G., Woolley, A.G., Miller, L.D., Black, M.A., et al. (2012). YB-1, the E2F pathway, and regulation of tumor cell growth. *J. Natl. Cancer Inst.* *104*, 133–146.
- Lee, Y.S., Shibata, Y., Malhotra, A., and Dutta, A. (2009). A novel class of small RNAs: tRNA-derived RNA fragments (tRFs). *Genes & Development* *23*, 2639–2649.
- Lujambio, A., and Lowe, S.W. (2012). The microcosmos of cancer. *Nature* *482*, 347–355.
- Matsumoto, K., and Bay, B.H. (2005). Significance of the Y-box proteins in human cancers. *J. Mol. Genet. Med.* *1*, 11–17.
- Minn, A.J., Gupta, G.P., Siegel, P.M., Bos, P.D., Shu, W., Giri, D.D., Viale, A., Olshen, A.B., Gerald, W.L., and Massague, J. (2005). Genes that mediate breast cancer metastasis to lung. *Nature* *436*, 518–524.
- Moyer, M.W. (2012). Targeting hypoxia brings breath of fresh air to cancer therapy. *Nat. Med.* *18*, 636–637.
- Salmena, L., Poliseno, L., Tay, Y., Kats, L., and Pandolfi, P.P. (2011). A ceRNA hypothesis: the Rosetta Stone of a hidden RNA language? *Cell* *146*, 353–358.
- Semenza, G.L. (1999). Regulation of mammalian O₂ homeostasis by hypoxia-inducible factor 1. *Annu. Rev. Cell Dev. Biol.* *15*, 551–578.
- Shen, C., and Kaelin, W.G., Jr. (2013). The VHL/HIF axis in clear cell renal carcinoma. *Semin. Cancer Biol.* *23*, 18–25.
- Speer, J., Gehrke, C.W., Kuo, K.C., Waalkes, T.P., and Borek, E. (1979). tRNA breakdown products as markers for cancer. *Cancer* *44*, 2120–2123.
- Thompson, D.M., and Parker, R. (2009). Stressing out over tRNA cleavage. *Cell* *138*, 215–219.
- Uchiiumi, T., Fotovati, A., Sasaguri, T., Shibahara, K., Shimada, T., Fukuda, T., Nakamura, T., Izumi, H., Tsuzuki, T., Kuwano, M., and Kohno, K. (2006). YB-1 is important for an early stage embryonic development: neural tube formation and cell proliferation. *J. Biol. Chem.* *281*, 40440–40449.
- Ule, J., Jensen, K., Mele, A., and Darnell, R.B. (2005). CLIP: a method for identifying protein-RNA interaction sites in living cells. *Methods* *37*, 376–386.
- Wilson, W.R., and Hay, M.P. (2011). Targeting hypoxia in cancer therapy. *Nat. Rev. Cancer* *11*, 393–410.
- Wu, Y., Yamada, S., Izumi, H., Li, Z., Shimajiri, S., Wang, K.Y., Liu, Y.P., Kohno, K., and Sasaguri, Y. (2012). Strong YB-1 expression is associated with liver metastasis progression and predicts shorter disease-free survival in advanced gastric cancer. *J. Surg. Oncol.* *105*, 724–730.
- Zhang, C.L., and Darnell, R.B. (2011). Mapping in vivo protein-RNA interactions at single-nucleotide resolution from HITS-CLIP data. *Nat. Biotechnol.* *29*, 607–614.

Neuronal Activity Promotes Glioma Growth through Neuroligin-3 Secretion

Graphical Abstract



Authors

Humsa S. Venkatesh, Tessa B. Johung, ..., Parag Mallick, Michelle Monje

Correspondence

mmonje@stanford.edu

In Brief

Neuronal activity promotes the growth of malignant glioma through activity-regulated secretion of the synaptic protein neuroligin-3, which acts as a mitogen, recruiting the PI3K-mTOR pathway to induce glioma cell proliferation.

Highlights

- Neuronal activity promotes high-grade glioma (HGG) proliferation and growth
- Neuroligin-3 is an activity-regulated secreted glioma mitogen
- Neuroligin-3 induces PI3K-mTOR signaling in HGG cells
- Neuroligin-3 expression is inversely correlated with survival in human HGG

Accession Numbers

GSE62563



Neuronal Activity Promotes Glioma Growth through Neuroligin-3 Secretion

Humsa S. Venkatesh,^{1,2,3,4,5,9} Tessa B. Johung,^{1,2,3,4,5,9} Viola Caretti,^{1,2,3,4,5,9} Alyssa Noll,^{1,2,3,4,5} Yujie Tang,^{1,2,3,4,5} Surya Nagaraja,^{1,2,3,4,5} Erin M. Gibson,^{1,2,3,4,5} Christopher W. Mount,^{1,2,3,4,5} Jai Polepalli,⁶ Siddhartha S. Mitra,⁵ Pamelyn J. Woo,^{1,2,3,4,5} Robert C. Malenka,⁶ Hannes Vogel,^{1,2,3,4} Markus Bredel,⁷ Parag Mallick,⁸ and Michelle Monje^{1,2,3,4,5,*}

¹Department of Neurology

²Department of Pediatrics

³Department of Pathology

⁴Department of Neurosurgery

⁵Institute for Stem Cell Biology and Regenerative Medicine

⁶Nancy Pritzker Laboratory, Department of Psychiatry and Behavioral Sciences
Stanford University School of Medicine, Stanford, CA 94305, USA

⁷Department of Radiation Oncology, University of Alabama at Birmingham School of Medicine, Birmingham, AL 35233, USA

⁸Department of Radiology, Stanford University School of Medicine, Stanford, CA 94305, USA

⁹Co-first author

*Correspondence: mmonje@stanford.edu

<http://dx.doi.org/10.1016/j.cell.2015.04.012>

SUMMARY

Active neurons exert a mitogenic effect on normal neural precursor and oligodendroglial precursor cells, the putative cellular origins of high-grade glioma (HGG). By using optogenetic control of cortical neuronal activity in a patient-derived pediatric glioblastoma xenograft model, we demonstrate that active neurons similarly promote HGG proliferation and growth in vivo. Conditioned medium from optogenetically stimulated cortical slices promoted proliferation of pediatric and adult patient-derived HGG cultures, indicating secretion of activity-regulated mitogen(s). The synaptic protein neuroligin-3 (NLGN3) was identified as the leading candidate mitogen, and soluble NLGN3 was sufficient and necessary to promote robust HGG cell proliferation. NLGN3 induced PI3K-mTOR pathway activity and feedforward expression of *NLGN3* in glioma cells. *NLGN3* expression levels in human HGG negatively correlated with patient overall survival. These findings indicate the important role of active neurons in the brain tumor microenvironment and identify secreted NLGN3 as an unexpected mechanism promoting neuronal activity-regulated cancer growth.

INTRODUCTION

High-grade gliomas (HGG), the leading cause of brain tumor death in both children and adults, occur in a striking spatiotemporal pattern highlighting the critical importance of the tumor microenvironment. Molecularly defined subtypes of HGG parse by neuroanatomical site of origin and patient age, with pontine

and thalamic HGGs typically occurring in mid-childhood, cortical gliomas of childhood occurring in older children and young adults, and HGG of later adulthood occurring chiefly in the frontotemporal lobes (Khuong-Quang et al., 2012; Schwartzentruber et al., 2012; Sturm et al., 2012; Wu et al., 2012). These age and neuroanatomical predilections of gliomagenesis point to interactions between cell of origin and microenvironment, suggesting dysregulation of neurodevelopment and/or plasticity.

Microenvironmental determinants of glioma cell behavior are incompletely understood, although important relationships between glioma cells and neighboring microglia, astrocytes, and vascular cells have recently come to light (Charles et al., 2011; Pyonteck et al., 2013; Silver et al., 2013). While cellular origins of HGG remain unclear, converging evidence implicates oligodendroglial precursor cells (OPCs) and earlier neural precursor cells (NPCs) as putative cells of origin for many forms of HGG (Galvao et al., 2014; Liu et al., 2011; Monje et al., 2011; Wang et al., 2009). Clues to microenvironmental influences driving HGG growth may thus be inferred from mechanisms governing the proliferation of normal NPCs and OPCs in the postnatal brain. We recently demonstrated that neuronal activity exerts a strong mitogenic effect on normal NPCs and OPCs in juvenile and adult mammalian brains (Gibson et al., 2014), raising the possibility that neuronal activity could promote proliferation in HGG.

RESULTS

Optogenetic Control of Cortical Neuronal Activity in a Patient-Derived Pediatric Cortical HGG Orthotopic Xenograft Model

To test the role of neuronal activity in HGG growth, we employed in vivo optogenetic stimulation of premotor cortex in freely behaving mice bearing patient-derived orthotopic xenografts of pediatric cortical glioblastoma (GBM; Figure 1A–1C). The well-characterized Thy1::ChR2 mouse model expressing the excitatory opsin channelrhodopsin-2 (ChR2) in deep cortical

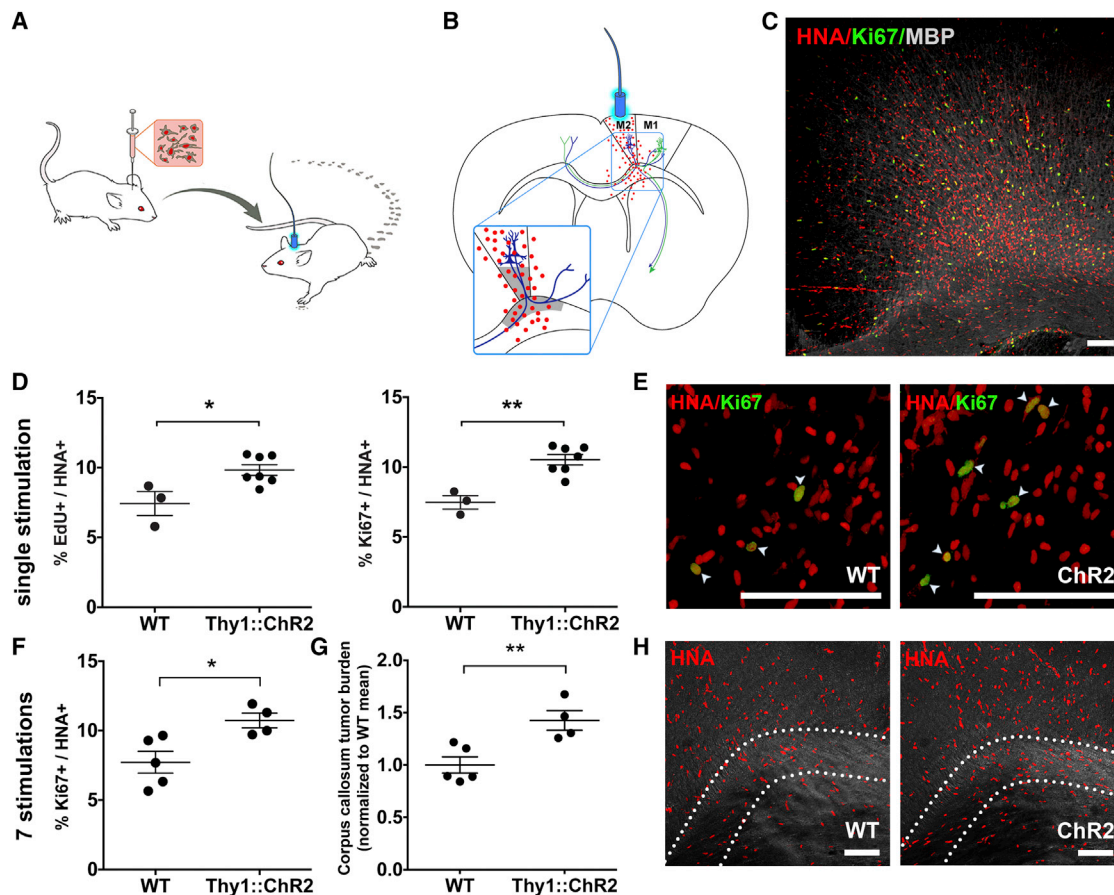


Figure 1. Neuronal Activity Promotes High-Grade Glioma Proliferation and Growth In Vivo

(A) In vivo optogenetic high-grade glioma (HGG) orthotopic xenograft model. (B) Schematic illustration of the optogenetically stimutable premotor circuit. Thy1::Chr2⁺ premotor cortex (M2) neurons depicted in blue; primary motor cortex (M1) projection neurons, green; tumor cells depicted as red dots. Gray shading indicates region of analysis. (C) Confocal micrograph of infiltrating pHGG (SU-pcGBM2) cells expressing human nuclear antigen (HNA, red), proliferation marker Ki67 (green) in premotor cortical deep layers and subjacent corpus callosum (MBP, white). (D and E) Single optogenetic stimulation session paradigm. (D) Proliferation index of pHGG cells in identically manipulated WT;NSG (n = 3) and Thy1::Chr2;NSG (n = 7) mice, measured by the proportion of HNA⁺ cells expressing EdU (left graph) or Ki67 (right graph) 24 hr after one optogenetic stimulation session. (E) Confocal micrograph illustrating proliferating (Ki67⁺, green) pHGG cells (HNA⁺, red) xenografted in WT;NSG ("WT"; left) or Thy1::Chr2;NSG mice ("Chr2"; right). (F–H) Repetitive optogenetic stimulation sessions paradigm. Xenografted WT;NSG (n = 5) and Thy1::Chr2;NSG (n = 4) mice evaluated 48 hr after seven daily sessions of optogenetic stimulation. (F) Proliferation index (Ki67⁺/HNA⁺) as in (D) above after seven stimulations. (G) Tumor cell burden increases following 1 week of brief daily optogenetic stimulation sessions, measured as HNA⁺ cell density within the region of corpus callosum containing active premotor projections; data normalized to WT mean. (H) Confocal micrographs with differential interference contrast (DIC) background to illustrate regional tissue architecture; HNA⁺ pHGG cells (red) are infiltrating premotor cortex and subjacent corpus callosum. Dotted line indicates region of analysis in corpus callosum. Data shown as mean ± SEM. *p < 0.05, **p < 0.01 by unpaired two-tailed Student's t test. Scale bars, 100 μm. See also [Figure S1](#) and [Movie S1](#).

projection neurons (Arenkiel et al., 2007; Wang et al., 2007) was crossed onto an immunodeficient background (NOD-SCID-IL2R γ-chain-deficient, NSG), resulting in a mouse model (Thy1::Chr2;NSG) amenable to both in vivo optogenetics and orthotopic xenografting. Chr2-expressing neurons respond with action potentials to 473 nm light pulses with millisecond precision (Arenkiel et al., 2007; Boyden et al., 2005; Wang et al., 2007). Expression of Chr2 does not alter membrane properties in the absence of light or neuronal health in the absence or presence of light under established experimental conditions (Boyden et al., 2005). When an optical fiber is placed just below the pial surface (Figure 1B), ~10% of the irradiance penetrates midway

through cortex, thus stimulating the apical dendrites of deep cortical projection neurons expressing Chr2 (Yizhar et al., 2011). Stimulating the premotor circuit unilaterally at 20 Hz, consistent with the 10–40 Hz physiological firing rate of motor cortex projection neurons, elicits complex motor behavior (unidirectional ambulation; Arenkiel et al., 2007; Gibson et al., 2014; Wang et al., 2007). Optogenetic stimulation of the premotor circuit elicits a substantial increase in NPC and OPC proliferation (Gibson et al., 2014). At baseline, precursor cell proliferation is equivalent in mice expressing or lacking Chr2 (Gibson et al., 2014). In this experimental paradigm, the microglial inflammatory response to superficial fiber placement and subsequent light

stimulation is minimal in deep cortex, where ChR2-expressing neurons reside, resolves within days, and is equal in Thy1::ChR2 mice and identically manipulated wild-type (WT) controls (Gibson et al., 2014).

To develop an orthotopic xenograft model appropriate to the juvenile premotor cortex, a culture was established from pre-treatment biopsy tissue of a frontal cortex GBM (WHO grade IV) from a 15-year-old male (culture designated SU-pcGBM2; clinical characteristics, genomic characterization and DNA fingerprinting described in Table S1). These pediatric cortical HGG (pHGG) cells were xenografted unilaterally into premotor (M2) cortex of juvenile Thy1::ChR2;NSG mice, resulting in diffusely infiltrating glioma cells throughout premotor cortex and subjacent corpus callosum (Figure 1C). WT (no opsin) littermate control NSG mice (WT;NSG) were identically manipulated for comparison. After tumors were allowed to engraft for 2 months, an optical-neural interface was placed ipsilateral to the xenograft. The unilateral premotor cortex was optogenetically stimulated (473 nm, 20 Hz; cycles of 30 s on/90 s off over 30 min) in awake mice, resulting in unidirectional ambulation. pHGG xenografts did not impede the behavioral response to evoked premotor circuit activity (Movie S1). Light stimulation had no behavioral effect in identically manipulated xenografted WT;NSG mice. Mice were given one dose of EdU to label proliferating cells at the time of optogenetic manipulation and were sacrificed 24 hr later to examine acute effects of neural activity.

Neuronal Activity Promotes High-Grade Glioma Growth In Vivo

Tumor cell burden and distribution did not differ between groups at the time of stimulation ($p = 0.74$; Figures S1A and S1B). Xenografted human tumor cells (human nuclear antigen, HNA⁺) co-expressing EdU indicate glioma cells proliferating at the time of EdU administration and optogenetic stimulation, while co-expression of Ki67 indicates cells proliferating at the time of sacrifice 24 hr later. Within the premotor circuit (Figures 1B and S1C), the proliferation index of human tumor cells was increased in optogenetically stimulated Thy1::ChR2;NSG mice compared to that of identically manipulated WT mice, measured both as the percent of tumor cells co-expressing EdU ($9.83\% \pm 0.38\%$ versus $7.43\% \pm 0.86\%$; $n = 7$ Thy1::ChR2;NSG mice, $n = 3$ WT;NSG mice, respectively; $p < 0.05$; Figures 1D and 1E) or co-expressing Ki67 ($10.53\% \pm 0.37\%$ versus $7.48\% \pm 0.48\%$; $n = 7$ Thy1::ChR2;NSG mice, $n = 3$ WT;NSG mice, respectively; $p < 0.01$, Figures 1D and 1E). This range of observed proliferation indices is consistent with that of human glioma. Proliferation index is typically <5% for low-grade astrocytomas, 5%–15% for anaplastic astrocytomas (WHO grade III), and 10%–20% for GBMs (WHO grade IV); proliferation indices correlate inversely with prognosis, with those above 10% generally indicating poor prognosis (Johannessen and Torp, 2006). The observed activity-regulated increase in proliferation was restricted to the active circuit; in the prefrontal cortex, a region infiltrated by glioma cells but outside of the area stimulated by light, glioma cell proliferation indices were equivalent in Thy1::ChR2;NSG and WT;NSG mice (Figures S1D and S1E). While proliferation increased within the active circuit, glioma cell death remained

constant, with only rare tumor cells expressing cleaved caspase-3 in either group (Figures S1F and S1G).

A simplified Galton-Watson mathematical model (Gerlee, 2013) of tumor cell growth incorporating the neuronal activity-associated increase in proliferation index (b) and a fixed cell death rate (d) would predict an exponential growth effect of elevated neuronal activity within the active circuit ($x_t = x_0(1+(b-d))^t$). Such a model utilizing the observed proliferation indices predicts an activity-regulated ~25% increase in tumor cell number after seven cell divisions and ~50% tumor increase after 14 divisions. To test this prediction in vivo, we utilized a repetitive stimulation paradigm in which mice were optogenetically manipulated as above for 10 min daily on 7 consecutive days and were sacrificed 48 hr after the final session. Following repetitive elevations in premotor circuit activity, tumor cell proliferation index was increased in xenografted Thy1::ChR2;NSG mice to a similar degree as in the single optogenetic stimulation paradigm ($10.74\% \pm 0.61$ versus $7.72\% \pm 0.88$; $n = 4$ Thy1::ChR2;NSG mice, $n = 5$ WT;NSG mice; $p < 0.05$, Figure 1F). As predicted, periodically elevated neuronal activity for 1 week yielded a ~42% increase in tumor cell burden within the active circuit relative to identically manipulated WT controls ($n = 4$ Thy1::ChR2;NSG, $n = 5$ WT;NSG mice; $p < 0.01$; Figures 1G and 1H). These data reflect the influence of neuronal activity on tumor burden during the exponential growth phase; over the course of the disease, as disruption of healthy tissue progresses and the microenvironment evolves, the effects of neuronal activity on glioma growth could change.

Neuronal Activity Promotes Glioma Proliferation through Secreted Factors

To determine whether neurons stimulate glioma proliferation via secretion of an activity-regulated mitogen(s), we optogenetically stimulated acute cortical slices from Thy1::ChR2 or identically manipulated WT mice in situ and collected the conditioned medium (CM), to which we exposed patient-derived HGG cultures (Figure 2A). The slice stimulation paradigm mirrored the in vivo paradigm, using 473 nm light at 20 Hz for cycles of 30 s on, 90 s off over a 30 min period. For this in situ optogenetic model, expected neuronal firing in response to light was validated electrophysiologically, confirming 20 Hz spike trains for 30 s periods throughout the 30 min session (Figure 2B). Maintenance of slice health throughout this paradigm was confirmed electrophysiologically and histologically (Figures S2A–S2D). Cortical slices from WT mice were identically manipulated for comparison. In parallel, CM was collected from blue light-unexposed Thy1::ChR2 and WT cortical slices. Patient-derived HGG cultures were then placed in CM from stimulated (light-exposed) or unstimulated cortical slices (Figure 2A). The HGG cell proliferation index (fraction of total cells in S phase as detected by EdU incorporation) was determined after 24 hr exposure to CM from the acute cortical slice conditions described above. The CM from optogenetically stimulated Thy1::ChR2 cortical slices (active CM) increased the in vitro proliferation index of pHGG (SU-pcGBM2) cells in comparison to CM from all control conditions, including identically manipulated WT, unstimulated Thy1::ChR2, or unstimulated WT cortical slices, or in comparison to blue light-exposed or non-exposed aCSF medium lacking

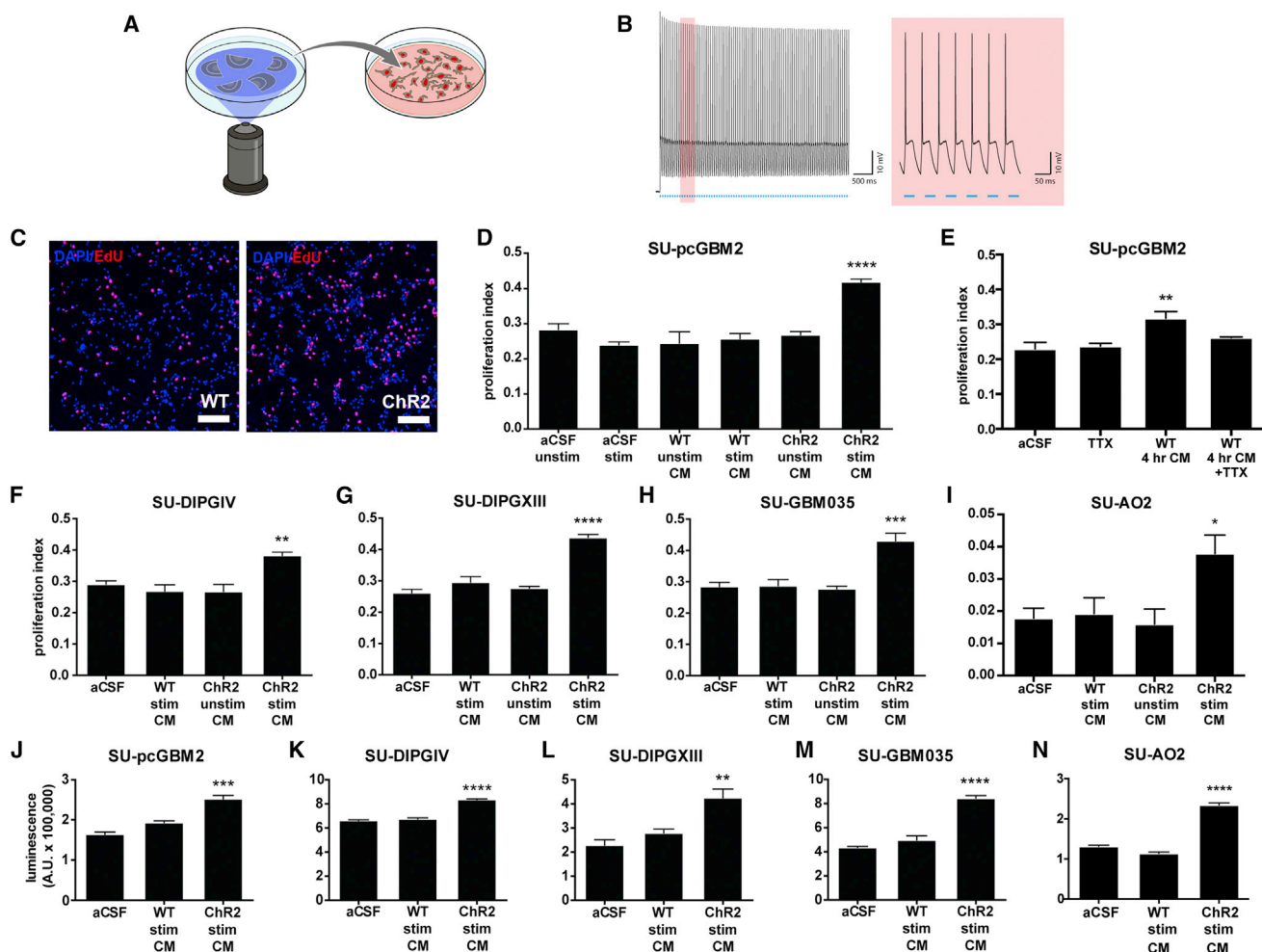


Figure 2. Activity-Regulated Secreted Factors Promote Glioma Cell Proliferation

(A) Schematic depicts optogenetic stimulation of acute cortical slices and collection of conditioned medium (CM).

(B) Electrophysiological demonstration by patch-clamp recording (left; trace highlighted in red is magnified at right) of 20 Hz neuronal firing in response to 20 Hz blue light pulses throughout the 30 s stimulation period in the Thy1::ChR2 cortical slice.

(C) Representative confocal micrographs show increased uptake of EdU (red) in cells (DAPI, blue) exposed to CM from stimulated Thy1::ChR2 slices (active CM) versus those exposed to CM from blue light-exposed WT slices (WT CM).

(D) Proliferation index of SU-pcGBM2 cells exposed to optogenetically stimulated or unstimulated Thy1::ChR2 cortical slice CM, blue light-exposed WT cortical slice CM ("WT stim CM") or non-exposed WT cortical slice CM ("WT unstim CM"), or plain media (aCSF).

(E) Proliferation index of SU-pcGBM2 cells after exposure to CM generated from light-unexposed WT slice conditioning for 4 hr in the presence or absence of 1 μ M tetrodotoxin (TTX).

(F–I) Active CM similarly increased the proliferation index of DIPG (F and G), adult GBM (H), and anaplastic oligodendroglioma (I) cultures.

(J–N) Active CM increased the viable cell number measured by CellTiter-Glo after 72 hr of incubation with active or light-exposed WT CM in pediatric and adult GBM (J and M), DIPG (K and L), and anaplastic oligodendroglioma (N) cells.

All experiments analyzed by one-way ANOVA and performed with $n = 3$ biological replicates. Data shown as mean \pm SEM. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$. Scale bars, 100 μ m. See also Figure S2 and Table S1.

slices ($F = 15.49$, $p < 0.0001$; Figures 2C and 2D). Active CM did not alter glioma cell death, as assessed by Annexin V FACS analysis (Figure S2E). The secretion of activity-regulated mitogen(s) was not frequency dependent, as CM from Thy1::ChR2 cortical slices optogenetically stimulated at 5 Hz elicited the same proliferative effect on pHGG cells (Figure S2F).

WT cortical slices do exhibit spontaneous neuronal activity; thus, we expect activity-regulated secreted factors to be present in WT CM, albeit to a lesser extent than in media conditioned by

Thy1::ChR2 slices with optogenetically elevated neuronal activity. To further explore the possible effects of spontaneous activity, we allowed WT slices to condition the media without blue light for 4 hr rather than 30 min in the presence or absence of the specific voltage-gated sodium channel blocker tetrodotoxin (TTX) to silence spontaneous action potentials. WT CM conditioned for a longer duration elicited an increase in pHGG proliferation; this effect was blocked in CM from slices incubated with TTX (proliferation index 0.32 ± 0.03 with 4 hr WT CM exposure

versus ~ 0.25 with aCSF, aCSF + TTX, or WT CM + TTX exposure; $F = 8.45$; $p < 0.01$; Figure 2E). Together, these data indicate that spontaneous neuronal activity regulates secretion of a glioma mitogen(s).

To determine whether this proliferative response to activity-regulated secreted factor(s) was specific to the pHGG model (SU-pcGBM2 cells) or more broadly applicable, we tested nine additional patient-derived HGG cell cultures (Table S1). All four tested cultures of diffuse intrinsic pontine glioma (DIPG), the most common form of pediatric HGG, demonstrated a similarly robust proliferative response to active CM exposure (Figures 2F, 2G, S2G, and S2H). We next tested four patient-derived adult GBM cultures and found a similar increase in cell proliferation after exposure to active CM (Figures 2H, S2I, and S2J) in all but one, which was a young adult epithelioid *BRAF*^{V600E} mutant GBM (SU-GBM047; Figure S2K). As the mitogenic effect appears largely generalizable across distinct HGG classes, we also tested a patient-derived culture of adult anaplastic oligodendroglioma and similarly observed increased cell proliferation in response to activity-regulated secreted factors (Figure 2I). Consistent with spontaneous neuronal activity of WT slices, some cultures exhibit a small but significant increase in proliferation in response to WT CM (e.g., SU-DIPGVI, proliferation index 0.21 ± 0.01 in aCSF versus 0.31 ± 0.02 in WT CM versus 0.46 ± 0.002 in active CM; Figures S2G and S2J).

To ascertain whether the observed effect indicated an increase in overall glioma growth, we used the quantitative viable cell assay CellTiter-Glo following 72 hr exposure to cortical slice CM and found an increase in viable HGG cell number when cultures were exposed to active CM (Figures 2J–2N and S2L–S2N).

Activity-Regulated Glioma Mitogen(s) Are Secreted Proteins

A series of biochemical analyses was employed to ascertain the nature of the activity-regulated mitogen(s). To determine whether the mitogen(s) are small molecules or macromolecules, active or control CM was collected as above and fractionated by molecular size. The >10 kDa macromolecular fraction of active CM, but not the <10 kDa fraction, increased the in vitro glioma proliferation index (Figure 3A). Subsequent fractionation indicated that the mitogen(s) were present in the <100 kDa fraction (Figure 3B). To determine the biochemical nature of the mitogen(s), active CM was heated to $>100^\circ\text{C}$ to denature proteins, resulting in loss of its mitogenic effect (Figure 3C). In contrast, treatment of active CM with RNase and DNase had no effect on its proliferation-inducing capacity (Figure 3D). Taken together, these data indicate that the neuronal activity-regulated secreted mitogen(s) is a protein between 10–100 kDa.

With respect to small molecules, high levels of glutamate release into CM would not be expected in a healthy brain slice, as perisynaptic astrocytes take up released glutamate from the synaptic cleft (Rothstein et al., 1996). Indeed, low levels of glutamate were present in active CM (Figure S3A). Thus, this in situ experimental paradigm does not address the potential role of glutamate in neuronal activity-regulated glioma cell proliferation in vivo.

Cortical Projection Neuronal Activity-Regulated Secretome

To identify the secreted protein(s) that increase glioma cell proliferation in an activity-dependent manner, we employed mass spectrometric analyses of the cortical slice CM. Of note, neuronal activity may regulate secretion of proteins from neurons themselves or from other cell types in response to active neurons. Active CM and light-exposed WT CM were analyzed and compared using 2D gel electrophoresis to separate the secreted proteins by size and charge; differentially secreted protein spots were then identified by mass spectrometry. The 2D gel analyses were performed in duplicate using independent samples (Figure 3E). Quantitative mass spectrometric techniques of spectral counting and tandem mass tags (TMT) were then used with a third set of independent samples to confirm the 2D gel results and to more precisely define the absolute and relative quantities of each protein in the CM samples. The intersection of these analyses most consistently and robustly identified neuroligin-3 (NLGN3) as the leading candidate mitogen (Figure 3F), present in active CM at a concentration of ~ 20 – 40 nM and upregulated by 2.6-fold compared to light-exposed WT CM. Additional candidates identified are listed in Figure 3G.

The neuroligins are a family of synaptic proteins with a large N-terminal ectodomain, single pass transmembrane domain, and smaller C-terminal cytoplasmic domain (Südhof, 2008). Neuroligin-1 (NLGN1), acting primarily at excitatory synapses similarly to NLGN3, is secreted in an activity-regulated fashion by enzymatic cleavage of the N-terminal ectodomain (Peixoto et al., 2012; Suzuki et al., 2012). The 2D gel and quantitative mass spectrometric analyses across all three independent samples demonstrated excellent coverage of the NLGN3 ectodomain amino acid sequence (protein prophet score = 1; Table S2), identifying the protein with high confidence (Figure 3H). However, the C-terminal transmembrane and cytoplasmic domain of the protein was not detected in NLGN3 isolated from the active CM (Figure 3H).

Secreted Neuroligin-3 Promotes Glioma Cell Proliferation

The sufficiency of NLGN3 to promote HGG cell proliferation was then tested in vitro. We obtained recombinant full-length human NLGN3 and confirmed its identity and purity by mass spectrometry (Figure S3B). In contrast to the NLGN3 present in the CM, mass spectrometric analysis of recombinant NLGN3 did identify peptide sequences within the C-terminal tail. 24 hr exposure of pHGG cells to recombinant NLGN3 at various concentrations in vitro promoted a significant increase in proliferation index (Figure 4A), with no change in cell death as measured by Annexin V FACS analysis (Figure 4B). NLGN3 at the concentration present in the active CM (20–40 nM) elicits an increase in proliferation commensurate with the effect of active CM (Figures 4A and 2D). NLGN3 promoted proliferation of each additional patient-derived cell culture tested, including DIPG, adult GBM, and anaplastic oligodendroglioma, with the exception of the epithelioid GBM culture (Figures 4C and S4A).

Additional candidates identified in the proteomic analyses above were also screened in pHGG cells in vitro. Of these, brain-derived neurotrophic factor (BDNF) and the known glioma

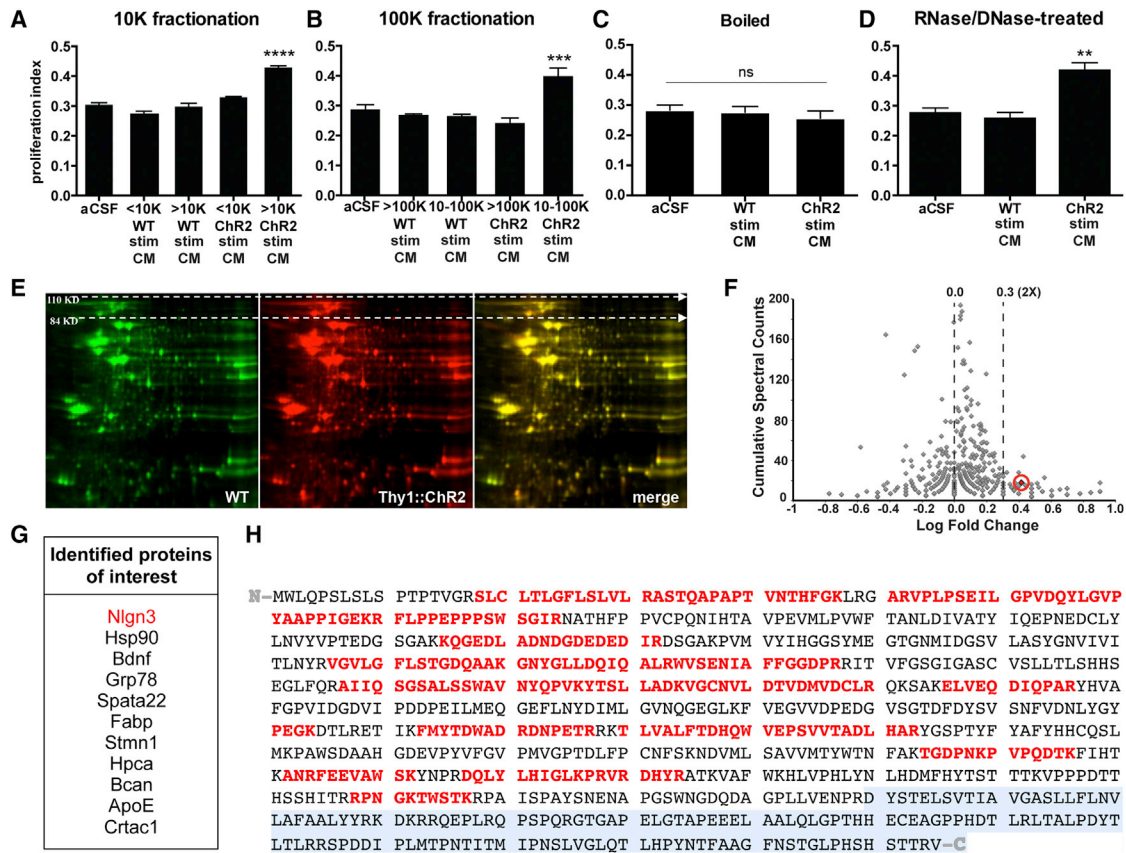


Figure 3. Cortical Neuronal Activity-Regulated Glioma Mitogen(s) Are Protein(s)

(A and B) Fractionation of CM by molecular size reveals that the activity-regulated mitogenic factors are >10 kDa (A) and <100 kDa (B). (C) Heating active CM to 100°C inactivates the mitogen(s). (D) RNA and DNA digestion of active CM does not change its mitogenic effect. All experiments analyzed by one-way ANOVA and performed with $n = 3$ biological replicates. Data shown as mean \pm SEM. ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$. n.s. indicates $p > 0.05$. (E) Representative two-dimensional gel electrophoresis separating proteins in light-exposed WT CM (green) and active CM (red) by size (vertical axis) and charge (horizontal axis); merged images, right-most panel. (F) Volcano plot of spectral counting data shows the ratio of peptides in a given protein found in active CM versus CM from unstimulated Thy1::ChR2 slices. Neuroligin-3 (Nlgn3) is highlighted and circled in red. (G) List of candidate proteins of interest identified from proteomic analyses. (H) Nlgn3 peptide sequence. Peptides in red were identified by mass spectrometry of the Nlgn3 isolated from active CM. Despite excellent coverage across the N-terminal ectodomain of the protein, no part of the C-terminal endodomain (transmembrane and intracellular domains, shaded gray) was identified in the isolated soluble Nlgn3. See also Figure S3 and Table S2.

mitogen 78 kDa glucose-regulated protein (GRP78; Lee et al., 2008) promoted pHGG proliferation but less potently than NLGN3 (Figures S4B–S4D). Additional candidates tested did not affect proliferation, even at high concentrations (Figure S4B). Thus, NLGN3 emerged as an unexpected cortical neuronal activity-regulated glioma mitogen, together with known mitogens BDNF and GRP78.

To test the necessity of Nlgn3 for the proliferation-promoting effect of active CM, we utilized the specific and avid binding of neuroligin-1 β (NRXN1 β) to NLGN3 (Ichtchenko et al., 1996) to deplete Nlgn3 from the cortical slice CM. Confirming that NRXN1 β in this setting does deplete the available Nlgn3, addition of NRXN1 β completely abrogated the mitogenic effect of recombinant NLGN3 exposure (proliferation index 0.40 ± 0.01 in pHGG

cells exposed to 50 nM NLGN3 versus 0.28 ± 0.01 in cells exposed to 50 nM NLGN3 + 500 nM NRXN1 β ; $p < 0.001$; Figure 4D). Addition of NRXN1 β alone to aCSF or to WT CM had no effect on proliferation index (Figure 4D). However, addition of NRXN1 β significantly decreased the mitogenic effect of active slice CM on pHGG cells (proliferation index 0.40 ± 0.01 in cells exposed to active CM versus 0.34 ± 0.01 with exposure to active CM + NRXN1 β ; $p < 0.05$; Figure 4D), indicating that secreted NLGN3 is necessary for the full mitogenic effect of cortical neuronal activity on glioma cells. The incomplete abrogation of the mitogenic effect of the CM with addition of NRXN1 β is consistent with additional activity-regulated glioma mitogens GRP78 and BDNF present in the CM (Figures S4B–S4D). Indeed, pharmacological inhibition of the BDNF receptor TRKB in

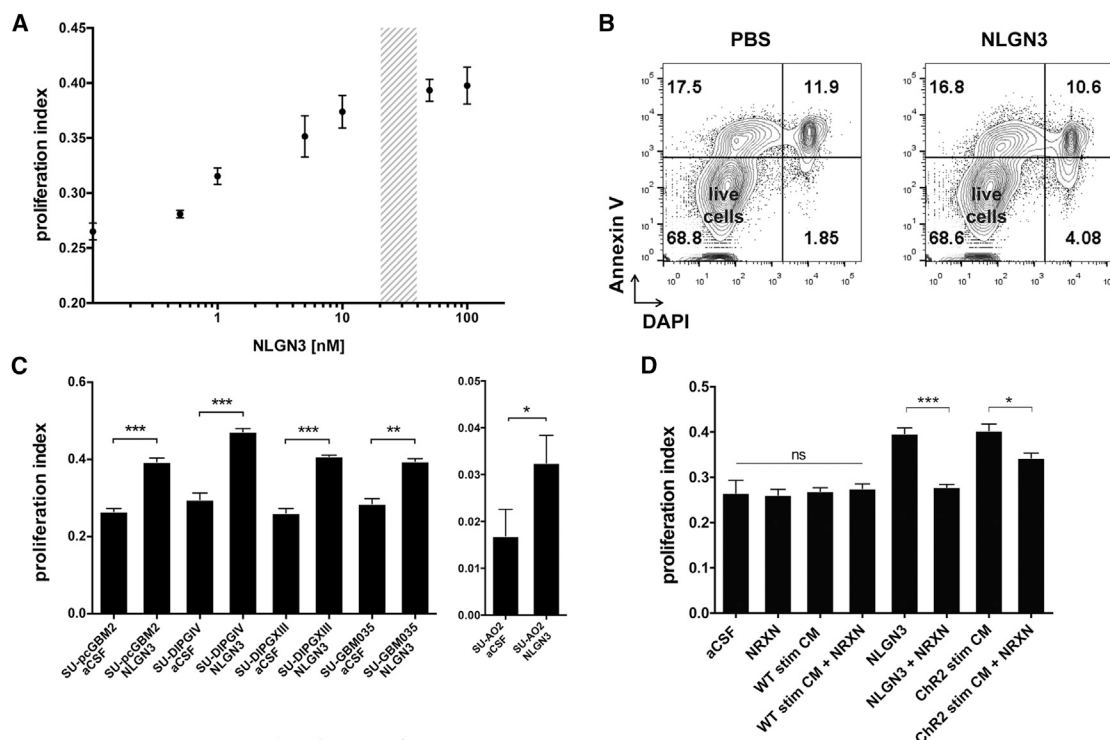


Figure 4. Secreted Neuroligin-3 Mediates Neuronal Activity-Regulated Glioma Proliferation

(A) Seven-point dose curve plots SU-pcGBM2 proliferation index as measured by EdU⁺/DAPI⁺ staining 24 hr after exposure to recombinant NLGN3 at a 0–100 nM concentration range. Shaded region indicates concentration present in active CM.

(B) After 24 hr exposure to PBS or NLGN3 (50 nM), SU-pcGBM2 cells were stained with DAPI (x axis) and Annexin V-FITC (y axis) to detect cell death by FACS analysis, performed in biological duplicate. Live Annexin V⁻/DAPI⁻ cells shown in lower-left quadrant of contour plots; pre-apoptotic Annexin V⁺/DAPI⁻ cells, left upper quadrant; dead Annexin V⁺/DAPI⁺ cells, right upper quadrant. No increase in cell death was seen with NLGN3 exposure.

(C) Proliferation indices of various patient-derived HGG cell lines exposed to 50 nM NLGN3 for 24 hr (unpaired two-tailed Student's t tests).

(D) Neurexin-1β (NRXN, 500 nM), which binds NLGN3 with high affinity, effectively blocks the mitogenic effect of recombinant NLGN3 (50 nM) and abrogates the mitogenic effect of active CM (unpaired two-tailed Student's t tests). Exposure to NRXN alone or added to light-exposed WT CM ("WT Stim CM") does not affect pHGG cell proliferation (one-way ANOVA).

For all experiments, n = 3 biological replicates unless otherwise noted. Data shown as mean ± SEM. *p < 0.05, **p < 0.01, ***p < 0.001. n.s. indicates p > 0.05. See also Figure S4.

combination with NRXN1β completely abrogated the proliferative effect of active CM (Figure S4E).

Downstream Mechanisms of Neuronal Activity-Regulated Glioma Proliferation

To begin to understand the intracellular signaling mechanisms by which neuronal activity promotes HGG cell proliferation, we performed RNA sequencing to define the transcriptome of pHGG cells exposed to active CM versus light-exposed WT CM. Pathway analysis of differentially regulated genes revealed upregulation of the immediate early gene and proto-oncogene *FOS*, whose expression can be downstream of pathways that include PI3K-mTOR signaling or MAPK signaling (Greenberg and Ziff, 1984; Gonzales and Bowden, 2002; Chen and Davis, 2003), suggesting potential involvement of either pathway (Table S3). Exposure to NLGN3 similarly resulted in upregulation of *FOS* expression, determined by qPCR (Figure 5A). However, western blot analysis did not reveal upregulation of phospho-ERK1/2^{T202/Y204}, an indicator of MAPK pathway activation, following NLGN3 exposure (Figures S5A and S5B). We thus

examined PI3K pathway recruitment by exposure to NLGN3 using western blot analysis of phospho-AKT^{S473}. pHGG cells exposed to NLGN3 exhibited increased phospho-AKT^{S473} levels relative to total AKT in a dose-dependent manner ($F = 17.99$, $p < 0.001$; Figure 5B-C). PI3K canonically regulates mammalian target of rapamycin (mTOR), and thus we examined the effect of NLGN3 exposure on mTOR activity using western blot analysis of phospho-4E-BP1^{T37/46}, revealing an increase in phospho-4E-BP1^{T37/46} relative to total 4E-BP1 following NLGN3 exposure (Figures 5D and 5E). Blockade of PI3K or mTOR pharmacologically or via shRNA-mediated knockdown prevented the NLGN3-mediated mitogenic effect (Figures 5F–5H and S5C–S5E). Neuronal activity-regulated secretion of NLGN3 thus recruits the PI3K-mTOR pathway to promote glioma cell proliferation.

Surprisingly, we also found upregulated expression of the neuroligin-3 gene (*NLGN3*) in pHGG cells exposed to active CM (Table S3), suggesting a positive feedforward effect on glioma cell *NLGN3* expression. To determine whether soluble NLGN3 exposure induces its own expression, we performed

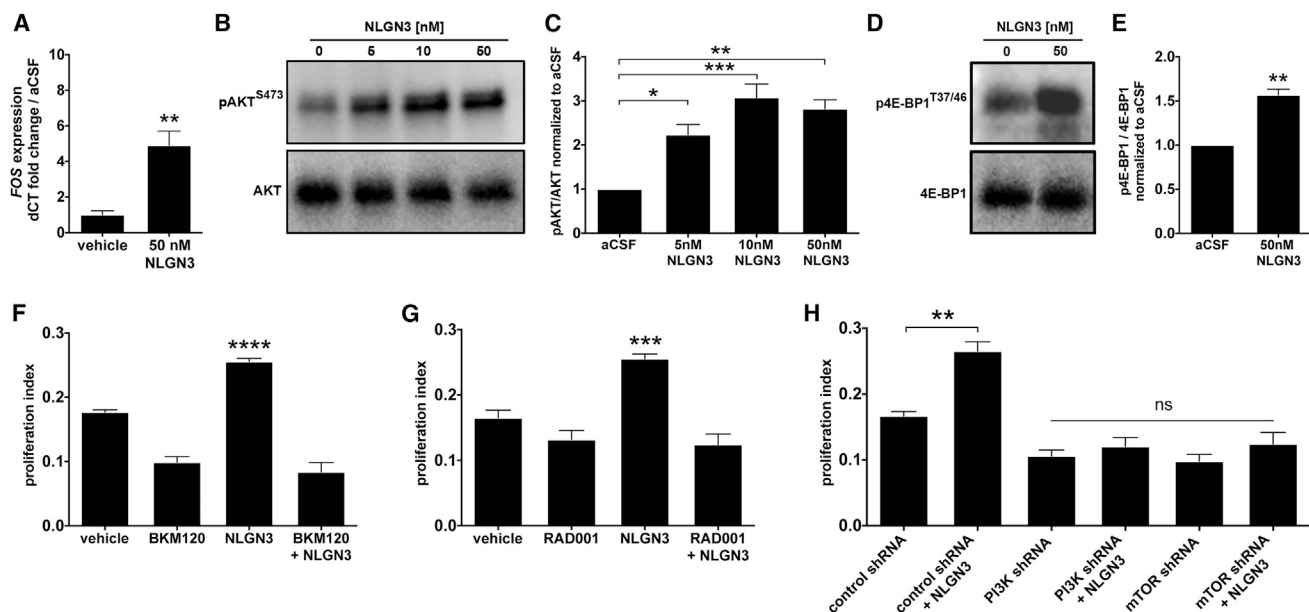


Figure 5. Secreted Neuroligin-3 Recruits the PI3K Pathway and Promotes Feedforward Expression of NLGN3

(A) FOS mRNA expression increases after 1 hr exposure to 50 nM NLGN3 compared to vehicle ($p < 0.01$ by unpaired two-tailed Student's *t* test). (B) NLGN3 increases PI3K pathway signaling. Representative western blot shows increased phosphorylation of AKT (pAKT^{S473}, top; total AKT, bottom) in response to NLGN3 concentrations ranging from 0 to 50 nM. (C) Quantification of the pAKT^{S473}/AKT ratio fold change (normalized to aCSF) observed in (B). (D) Representative western blot demonstrates increased phosphorylation of 4E-BP1, a downstream reporter of mTOR, after 50 nM NLGN3 exposure. (Top) 4E-BP1^{T37/46}, (bottom) total 4E-BP1. (E) Quantification of p4E-BP1^{T37/46}/4E-BP1 ratio fold change after NLGN3 exposure normalized to aCSF control (unpaired two-tailed Student's *t* test). (F) 50 nM NLGN3-induced increase in SU-pcGBM2 proliferation index (EdU assay) is blocked by inhibition of PI3K by BKM120 (100 nM). (G) Similar to (F), inhibition of mTOR by RAD001 (100 nM) blocks 50 nM NLGN3-induced proliferation in SU-pcGBM2 cells. (H) Genetic knockdown using specific shRNA against either *PI3K* or *mTOR* blocks effect of 50 nM NLGN3 on proliferation index (EdU assay in SU-pcGBM2). * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$, **** $p < 0.0001$ by one-way ANOVA with Tukey's post hoc tests to further examine pairwise comparisons unless otherwise indicated. All experiments performed in $n = 3$ biological replicates. Data shown as mean \pm SEM. See also Figure S5 and Table S3.

qPCR in cells exposed to recombinant NLGN3 and found that this elicits increased glioma cell *NLGN3* gene expression, tested in both pediatric cortical HGG ($F = 9.70$, $p < 0.01$; Figure 6A) and DIPG cells ($F = 13.56$, $p < 0.01$; Figure 6B). The role of PI3K-mTOR pathway activity in this positive feedforward effect was investigated using treatment with the PI3K inhibitor BKM120 or shRNA-mediated *PI3K* knockdown (Figures 6A–6D), both of which blocked the soluble NLGN3-induced increase in *NLGN3* gene expression (Figures 6A–6D and S6). Similarly, the mTOR inhibitor RAD001 or shRNA-mediated *mTOR* knockdown prevented the feedforward effect of NLGN3 on *NLGN3* gene expression (Figures 6C and 6E–6G). Soluble NLGN3 thus promotes glioma cell feedforward expression of *NLGN3* via the PI3K-mTOR pathway. To determine whether other ligands known to stimulate the PI3K pathway in glioma (Fan et al., 2009) similarly affect *NLGN3* expression, we tested the effect of epidermal growth factor (EGF) exposure on glioma cell *NLGN3* expression and found no effect ($p = 0.781$; Figure 6H), suggesting that *NLGN3* expression is specific to the context of NLGN3 exposure. Protein expression of NLGN3 following glioma cell NLGN3 exposure was confirmed using western blot analysis; in contrast, glioma cell NLGN3 protein expression was not found at baseline culture conditions (Figure 6I). NLGN3 thus results in feedforward

expression at the transcriptional and translational levels. Together, these findings indicate that *NLGN3* expression is an indicator of neuronal activity-dependent NLGN3 signaling to glioma cells (Figure 6J).

Neuroligin-3 Gene Expression Is Associated with Decreased Survival in Human High-Grade Glioma

Having demonstrated that NLGN3 exposure increases tumor cell *NLGN3* expression, we next asked whether the *NLGN3* gene exhibited aberrations in glioma. Analysis of data from The Cancer Genome Atlas (TCGA) showed that somatic mutations in *NLGN3* are infrequent in pediatric (pilocytic astrocytoma/medulloblastoma, 0.4%) and adult brain tumors (low-grade gliomas, 1.1%; high-grade gliomas, 0.4%; Table S4). Interestingly, an extended analysis of *NLGN3* mutations and copy-number aberrations across multiple cancer types in the International Cancer Genome Consortium (ICGC) and the cBioPortal for Cancer Genomics databases revealed more frequent mutations and amplifications in other tumors, with particular predominance in thyroid, pancreatic, prostate, and gastric cancers (Figure S7C and Table S4).

To validate the clinical significance of NLGN3 in human glioma pathophysiology, we next examined the relationship between

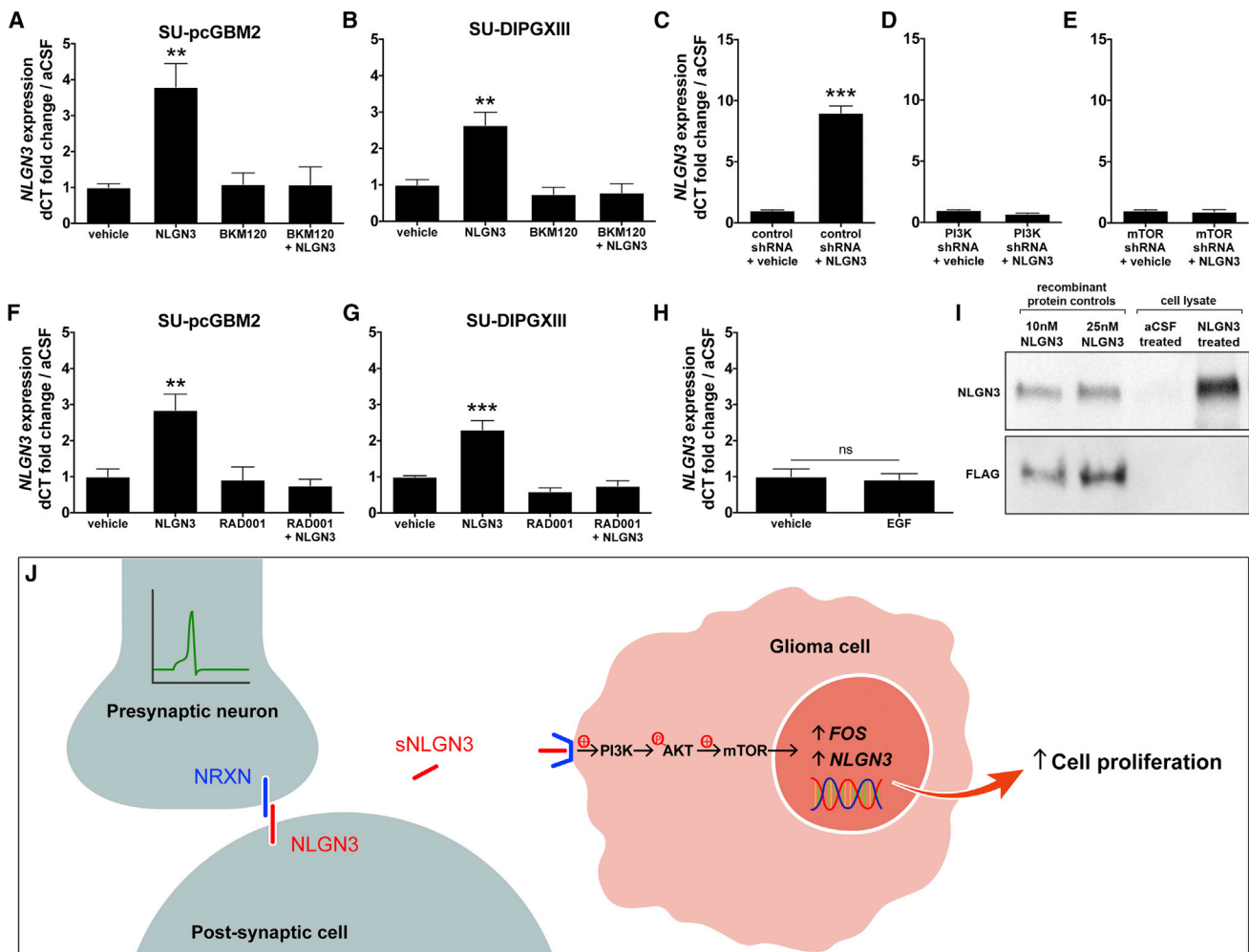


Figure 6. Secreted Neuroligin-3 Promotes Feedforward Expression of *NLGN3* through Recruitment of the PI3K-mTOR Pathway

(A) *NLGN3* mRNA expression in SU-pcGBM2 cells after 12 hr exposure to vehicle, 50 nM NLGN3, 100 nM BKM120, or 50 nM NLGN3 + 100 nM BKM120. (B) As in (A), SU-DIPGXIII *NLGN3* mRNA expression after exposure to NLGN3 and BKM120 alone or in combination. (C–E) *NLGN3* mRNA expression in SU-pcGBM2 cells after shRNA-mediated knockdown of either *PI3K* or *mTOR*. Only cells exposed to scrambled shRNA control exhibit increased *NLGN3* expression after NLGN3 exposure (unpaired two-tailed Student's *t* test.) (F) *NLGN3* mRNA expression in SU-pcGBM2 cells after 12 hr exposure to vehicle, 50 nM NLGN3, 100 nM RAD001, or 50 nM NLGN3 + 100 nM RAD001. (G) As in (F), SU-DIPGXIII *NLGN3* mRNA expression after exposure to NLGN3 and RAD001 alone or in combination. (H) *NLGN3* mRNA expression in SU-pcGBM2 cells does not change after 12 hr exposure to 50 nM EGF versus vehicle (unpaired two-tailed Student's *t* test). All qPCR data (A–H) are normalized to vehicle-treated samples and represent fold change of the delta CT in reference to β -actin. (I) Western blot analysis illustrating NLGN3 protein expression. Lanes 1 and 2 = 10 nM and 25 nM recombinant FLAG-tagged NLGN3, respectively. Lanes 3 and 4 = lysate from SU-pcGBM2 cells exposed to aCSF or 50 nM recombinant FLAG-tagged NLGN3, respectively. Top panel probed with anti-NLGN3; bottom panel probed with anti-FLAG. (J) Schematic illustrating the model of neuronal activity-regulated NLGN3 secretion from a post-synaptic cell, subsequent recruitment of glioma cell PI3K-mTOR pathway, expression of *FOS* and *NLGN3*, and proliferation. *n* = 3 biological replicates unless otherwise stated. Data shown as mean \pm SEM. **p* < 0.05, ***p* < 0.01, ****p* < 0.001 by one-way ANOVA unless otherwise stated. n.s. indicates *p* > 0.05. See also Figure S6 and Table S3.

NLGN3 gene expression and patient survival in 429 cases of adult GBM in TCGA. *NLGN3* mRNA expression level was found to be inversely correlated with patient overall survival (Figure 7). A two-class model in which patients were stratified according to median *NLGN3* expression showed an association between higher *NLGN3* expression and shorter survival (*p* < 0.05 by the log-rank test; Figure 7A). In patients whose tumors exhibited

below-median *NLGN3* expression, estimated mean survival was 20.8 months (95% CI, 17.1–24.4); in comparison, mean survival of patients with above-median *NLGN3* expression was 15.2 months (95% CI, 13.1–17.2). On Cox regression analysis, the hazard ratio for death with high versus low *NLGN3* expression was 1.31 (95% CI, 1.05–1.63). *NLGN3* expression was also significantly inversely associated with patient survival in a

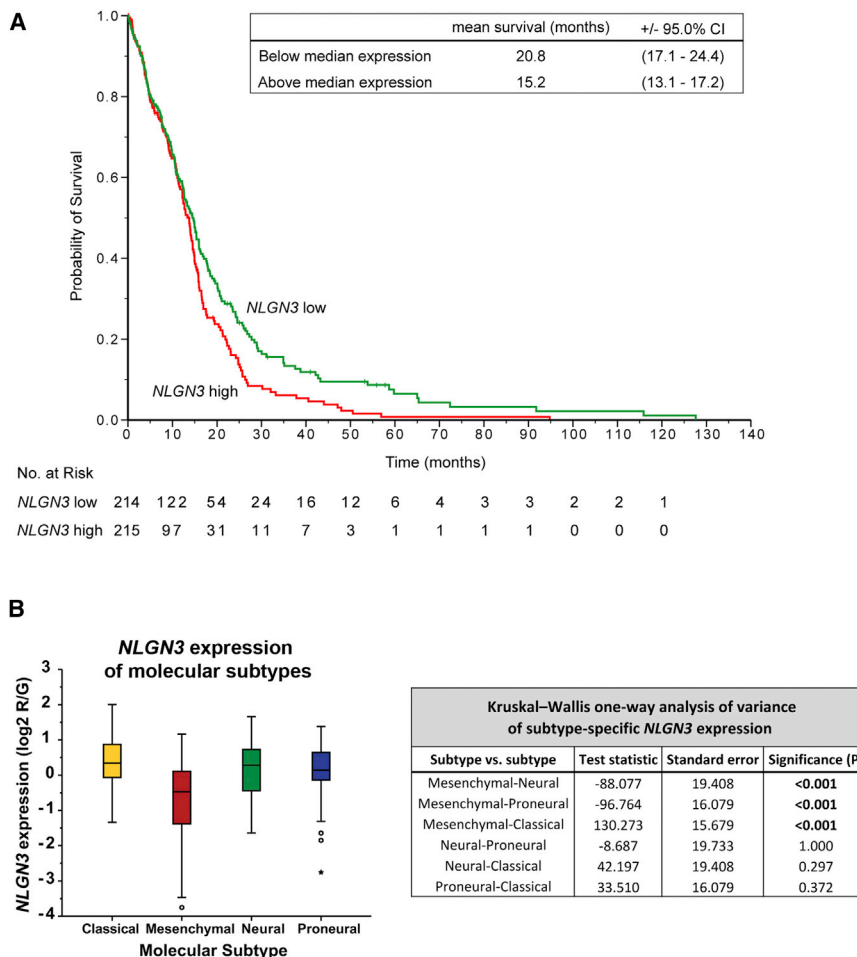


Figure 7. Neuroligin-3 Expression Inversely Correlates with Survival in Human Glioblastoma

(A) A two-class model stratified by median *NLGN3* expression in 429 GBM cases with molecular subtype data from the TCGA (<http://cancergenome.nih.gov>). Mean overall survival decreases by ~5.6 months in patients with tumors exhibiting above-median *NLGN3* expression; $p < 0.05$ by the log-rank test.

(B) GBM subtype-specific *NLGN3* expression. Box plots show the smallest and largest observations (top and bottom whiskers, respectively), the interquartile (IQ) range (box), and the median (black line). Data points more than 1.5 times the IQ range lower than the first quartile or 1.5 times the IQ range higher than the third quartile were considered outliers (shown as circles outside the box and whisker plot). Corresponding table of Kruskal-Wallis one-way ANOVAs with p values indicates pairwise comparisons of *NLGN3* expression in the four subtypes and significance of differential *NLGN3* expressions. See also Figure S7, Table S4.

DISCUSSION

Neurons in the Glioma Microenvironment

The results presented here demonstrate that excitatory neuronal activity can influence brain cancer growth. This represents a striking example of the core physiological function of an organ promoting the growth of a cancer arising within it.

An important mechanism mediating this

continuous Cox proportional-hazards regression analysis, such that higher expression represented an unfavorable prognosis (hazard ratio for death with high versus low *NLGN3* expression, 1.15; 95% CI, 1.01–1.30; $p < 0.05$).

To examine the specificity of these findings, we explored the relationship of neuroligin-2 (*NLGN2*) to survival in human GBM. Recombinant *NLGN2* does not promote pHGG proliferation in vitro (Figure S7D). Likewise, there is no significant association between *NLGN2* expression and patient survival in adult GBM, assessed as above in a continuous Cox model (hazard ratio for death with high versus low *NLGN2* expression, 0.95; 95% CI, 0.78–1.16; $p = 0.634$) and in a two-class model stratified by median expression ($p = 0.795$ by the log-rank test; Figure S7E).

Interestingly, upon examination of *NLGN3* expression by molecular GBM subtype as defined by TCGA (Verhaak et al., 2010), *NLGN3* expression was significantly lower in the mesenchymal subtype compared to classical, neural, and proneural subtypes (asymptotic significance of $p < 0.001$ by independent-samples Kruskal-Wallis test; Figure 7B). Notably, *NLGN3* expression remained significantly associated with patient survival in a multivariate Cox model that incorporates molecular subtype (hazard ratio for death with high versus low *NLGN3* expression, 1.15; 95% CI, 1.01–1.30; $p < 0.05$).

key microenvironmental interaction is activity-regulated secretion of *NLGN3*. The importance of *NLGN3* in HGG pathophysiology is underscored by the finding that *NLGN3* expression strongly predicts survival in human HGG. Taken together, these data elucidate a fundamental dimension of the HGG microenvironment and identify a robust and targetable mechanism driving HGG proliferation.

The role that neurons may play in brain cancer is underscored by perineuronal satellitosis, the histopathological hallmark of multiple forms of glioma characterized by tumor cell clustering around neuronal somata (Scherer 1938). A wealth of elegant data illustrates that neurotransmitters and neuropeptides can affect glioma cell behavior (Cuddapah et al., 2014; Labrakakis et al., 1998; Seifert and Sontheimer, 2014; Synowitz et al., 2001); for example, glutamate secreted from glioma cells influences their own proliferation and invasion through autocrine/paracrine signaling and subsequently increases the excitability of affected cortical networks (Buckingham et al., 2011; Campbell et al., 2012; Ishiuchi et al., 2002, 2007). However, a direct influence exerted by active parenchymal neurons upon the glioma environment has not been well appreciated. The critical role of neural elements in the cancer microenvironment has recently been elucidated for prostate (Magnon et al., 2013), pancreatic

(Stopczynski et al., 2014) and gastric (Zhao et al., 2014) cancers, in which peripheral innervation was found to potentially promote cancer progression. Our results suggest that active neurons play an important role in the microenvironment of brain tumors through malignant hijacking of mechanisms central to brain plasticity.

Normal and Malignant Neuron-Glial Interactions

A surprisingly broad range of molecularly and clinically distinct classes of HGG exhibited neuronal activity-regulated proliferation, and this response to neuronal activity mirrors that of putative HGG cells of origin. While the cellular origins of HGG may vary among subtypes of the disease, mounting evidence suggests that not only oligodendrogliomas (Persson et al., 2010; Sugiarto et al., 2011), but also many high-grade astroglomas, arise from precursor cells in the oligodendroglial lineage, including NPCs/early OPCs (Monje et al., 2011) and OPCs (Liu et al., 2011; Galvao et al., 2014; Glasgow et al., 2014). OPCs, the most mitotically active cells in the postnatal brain (Geha et al., 2010), may be particularly susceptible to malignant transformation due to Olig2-mediated suppression of p53 function (Mehta et al., 2011). Overexpression of a single transcription factor can determine the oligodendroglial or astrocytic phenotype of oligodendroglial lineage cell-derived tumors (Glasgow et al., 2014). A salient example of this point is the tumor from which the adult GBM culture SU-GBM052 used in this study was derived, a grade IV astrocytoma apparently arising from a transformed grade II oligodendroglioma. The “proneural,” and to a lesser extent the “neural” and “classical,” molecular subtypes of adult GBM, are defined by expression of oligodendroglial lineage-associated genes (Verhaak et al., 2010). Intriguingly, our data demonstrate that *NLGN3* is expressed abundantly in these three subtypes of GBM compared to “mesenchymal” GBM, supporting the concept of a lineage-specific molecular relevance of *NLGN3* to gliomagenesis. Normal NPCs and OPCs respond briskly to neuronal activity, and in the healthy juvenile and adult brain this response results in the activity-regulated generation of mature oligodendrocytes and remodeling of myelin, improving the function of that active circuit (Gibson et al., 2014). The findings presented here suggest that the malignant counterparts of these activity-responsive neural precursor cells may exploit mechanisms of myelin development and plasticity to promote growth.

Neuroigin-3 in Health and Disease

The finding that *NLGN3* is a glioma mitogen opens numerous doors to a deeper mechanistic understanding of its role in health and disease. The neuroilgins are post-synaptic adhesion molecules that are important in synaptic function and plasticity (Südhof, 2008; Varoqueaux et al., 2006). Nlgn1 and Nlgn3 are found in excitatory synapses, while Nlgn2 participates in inhibitory synapses (Gibson et al., 2009; Südhof, 2008). The canonical binding partners of the neuroilgins are presynaptic β -neurexins (Ichtchenko et al., 1996; Südhof, 2008). While wild-type neuroilgins play a central role in normal synaptic function, *NLGN3* mutations are implicated in altered synaptic function in autism (Jamain et al., 2003; Rothwell et al., 2014; Tabuchi et al., 2007). Our data show that somatic mutations and amplifications in

NLGN3 are also found at varying frequency in different types of human malignancies, supporting a possibly broader role of *NLGN3* in cancer. Such mutations are less frequent in gliomas, implying that non-genetic mechanisms of *NLGN3* deregulation may predominate in neoplasms of organs that normally express *NLGN3*. Such mechanisms may include, as we show here, activity-regulated secretion coupled with a positive feedforward effect on expression. Interestingly, *NLGN3* mutations and amplifications are prominent in pancreatic, prostate, and gastric cancers, for which a cancer growth-promoting role of innervation has been demonstrated (Stopczynski et al., 2014; Magnon et al., 2013; Zhao et al., 2014).

Neuroigin Secretion

Nlgn1 and Nlgn2 are known to undergo activity-dependent cleavage at the C-terminal transmembrane and cytoplasmic domain with resultant secretion of the N-terminal ectodomain (Peixoto et al., 2012; Suzuki et al., 2012). The data presented here illustrate activity-regulated secretion of Nlgn3 in the context of cortical projection neuronal activity. The mechanism of Nlgn3 secretion remains to be seen, but the apparent absence of the C-terminal transmembrane and cytoplasmic domain in the Nlgn3 protein identified in the active slice CM suggests similarities with mechanisms of Nlgn1 and Nlgn2 secretion. Given the enormous complexity of neurexin splice variants (> 1,000; Ullrich et al., 1995) and possible alternative binding partners (Samarelli et al., 2014), the identity of the binding partner for *NLGN3* in glioma cells remains an open question for further exploration.

While the present study provides evidence that active neurons promote HGG proliferation, this intercellular interaction may be indirect. Neuronal activity influences multiple cell types within an active neural circuit, and numerous activity-responsive cell types could play a role in promoting glioma growth. It is possible that *NLGN3* is secreted directly from active neurons or from OPCs, which act as post-synaptic cells in axoglial synapses (Bergles et al., 2010; Bergles et al., 2000) and express the highest level of *Nlgn3* mRNA of any neural cell type (Zhang et al., 2014).

Neuroigin-3, PI3K Pathway, and Feedforward Expression

How *NLGN3* recruits the glioma cell PI3K pathway is not yet clear, but links between neuroigin binding and receptor tyrosine kinase (RTK) activation, frequently upstream of PI3K, have been described in other contexts. Nlgn1 binding to Nrnx1 β in presynaptic neurons promotes neurite outgrowth in a manner that depends upon Nrnx1 β -mediated activation of the RTK fibroblast growth factor receptor-1 (Fgfr1; Gjølund et al., 2012). A similar NRXN-RTK interaction may mediate *NLGN3* stimulation of PI3K activity in HGG if a NRXN family member is indeed the binding partner of *NLGN3* in glioma cells.

PI3K-mediated feedforward regulation of *NLGN3* gene and protein expression was unexpected. *NLGN3* is not part of the canonical PI3K gene expression signature (Creighton et al., 2010), and, as demonstrated above, other growth factors known to stimulate PI3K such as EGF do not elicit changes in *NLGN3* gene expression in glioma cells, indicating that *NLGN3* is not a general marker of PI3K activity but is rather specific to this context. PI3K has been shown to regulate *Nlgn1* and *Nlgn2*

translation (Gkogkas et al., 2013), suggesting another link between PI3K pathway activity and regulation of *NLGN3* expression. Future work will elucidate the manner in which *NLGN3* recruits PI3K and subsequently promotes feedforward glioma *NLGN3* expression.

Additional Mechanisms of Neuronal Activity-Regulated Glioma Proliferation

It is also important to note that *NLGN3* is almost certainly not the only important mechanism promoting activity-regulated glioma growth. Indeed, GRP78 and BDNF were also identified as glioma mitogens, and accordingly, *Nlgn3* depletion resulted in a significant but incomplete abrogation of the mitogenic capacity of the CM, indicating partial contributions from other mitogens such as these. Recent single-cell analyses have elucidated intratumoral cellular heterogeneity in HGG (Patel et al., 2014), and it is not yet clear if these different activity-regulated mitogens act on the same or different cellular subpopulations. The candidate activity-dependent mitogens were recognized here via differential regulation and secretion that enabled their identification in the CM following a burst of cortical neuronal activity. Beyond these candidates, additional possible mitogens could be secreted in a more local manner or on a different timescale that precludes their identification within this experimental paradigm. Cell-contact-mediated mechanisms of activity-regulated glioma growth were also not evaluated in our in situ system. One mechanism that we do not evaluate explicitly is activity-regulated glutamate release. Certainly, glutamate released by glioma cells (Buckingham et al., 2011; Campbell et al., 2012; Ishiuchi et al., 2002, 2007) is well-demonstrated to promote glioma growth, and local neuronal glutamate release could function similarly, possibly contributing to the mitogenic effect of active neurons witnessed in vivo.

Conclusions

NLGN3 was identified as an unexpected mitogen promoting HGG growth. It is yet unclear whether *NLGN3* similarly mediates healthy myelin plasticity, which future work should address. If that proves to be the case, this important synaptic protein could represent a mechanism of coupling synaptic plasticity and myelin plasticity. Regardless, neuron-glioma cell interactions, including *NLGN3* secretion and subsequent signaling to the oncogenic PI3K-mTOR pathway in glioma cells, represent therapeutic targets for this group of devastating brain tumors.

EXPERIMENTAL PROCEDURES

See the [Extended Experimental Procedures](#) for detailed experimental procedures.

Isolation and Culture of the Primary Human Tumor Cells

Tumor tissue was dissociated and cultured as described in the [Extended Experimental Procedures](#).

Orthotopic Xenografting

600,000 SU-pcGBM2 cells were stereotactically implanted into the M2 pre-motor cortex of Thy1::ChR2;NSG or WT;NSG littermate mice at P35. Cells were allowed to engraft for at least 2 months prior to placement of an optical-neural interface for optogenetic stimulation.

In Vivo Optogenetic Stimulation

At least 7 days prior to stimulation, the optical-neural interface was placed just below the pial surface of the cortex ipsilateral to xenografts. For the single stimulation paradigm, animals were stimulated with cycles of 473 nm light pulses at 20 Hz for 30 s, followed by 90 s of recovery over a 30 min period and sacrificed 24 hr after stimulation. For the repetitive stimulation paradigm, animals were stimulated as above for 10 min periods on 7 consecutive days and were sacrificed 48 hr after the final session.

Generation of Conditioned Media

In short, Thy1::ChR2 or WT mouse brains were cut in 350 μ m sections on a vibratome, allowed to recover, and stimulated at 20 Hz using a blue-light LED transmitted through the microscope objective. Surrounding medium was then collected for immediate use or frozen at -80°C for future experiments.

Determination of Cell Proliferation In Vitro

To assess the number of cells actively entering S phase in response to various conditions (see [Extended Experimental Procedures](#)), patient-derived glioma cell cultures were exposed to cortical slice CM or various recombinant proteins along with 10 μ M EdU and were fixed after 24 hr. EdU incorporation was determined using Click-IT EdU visualization (Invitrogen).

Proteomic Analysis

Determination of proteins found within the active CM was done using 2D-gel electrophoresis accompanied by LC MS/MS.

Western Blot Analysis

Protein levels were determined using western blot analyses. Briefly, after various treatments, 400,000 SU-pcGBM2 cells were lysed and loaded onto SDS-PAGE gels. Proteins were separated with gel electrophoresis and transferred to a PVDF membrane. Proteins were then probed with various antibodies as described in [Extended Experimental Procedures](#).

qPCR

After various cell treatments (described in [Extended Experimental Procedures](#)), RNA was extracted from 500,000 SU-pcGBM2 or SU-DIPGXIII cells using TRIzol reagent. cDNA was generated using RT-PCR, and gene expression changes were further probed using quantitative PCR.

ACCESSION NUMBERS

The accession number for RNA-seq data deposited in the GEO database is GEO: GSE62563.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, one movie, and four tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.04.012>.

AUTHOR CONTRIBUTIONS

H.S.V. and T.B.J. conducted all experiments and contributed to experimental design, manuscript editing, and data analysis for all sections. V.C. contributed to conducting in vivo experiments and to experimental design. M.M. designed all experiments, wrote the manuscript, and supervised all aspects of the work.

ACKNOWLEDGMENTS

The authors gratefully acknowledge support from the Matthew Larson Foundation, Godfrey Family Fund in Memory of Fiona Penelope, National Institute of Neurological Disorders and Stroke (NINDS K08NS070926), McKenna Claire Foundation, National Science Foundation Graduate Research Fellowship Program, California Institute for Regenerative Medicine (CIRM RB4-06093 and RN3-06510), Alex's Lemonade Stand Foundation, The Cure Starts Now

Foundation, Lyla Nsouli Foundation, Unravel Pediatric Cancer, the Wayland Villars DIPG Foundation, the Dylan Jewett, Connor Johnson, Zoey Ganesh, Dylan Frick, Abigail Jensen, and Jennifer Kranz Memorial Funds, Virginia and D.K. Ludwig Fund for Cancer Research, the Bear Necessities Pediatric Cancer Foundation, Lucile Packard Foundation for Children's Health, Stanford University School of Medicine Dean's Fellowship, Child Health Research Institute at Stanford Anne T. and Robert M. Bass Endowed Faculty Scholarship in Pediatric Cancer and Blood Diseases, National Cancer Institute (PSOC-MCSTART U54CA143907), Canary Foundation, Ben and Catherine Ivy Foundation, and the Sujal and Meera Patel Foundation. Special thanks to Sarah Chen for illustrations, to Elizabeth Hoyte for assistance with figures, to Eric Raabe for use of the JHH-DIPGI cell line, and to Tom Südhof for helpful discussions.

Received: October 17, 2014

Revised: January 24, 2015

Accepted: March 3, 2015

Published: April 23, 2015

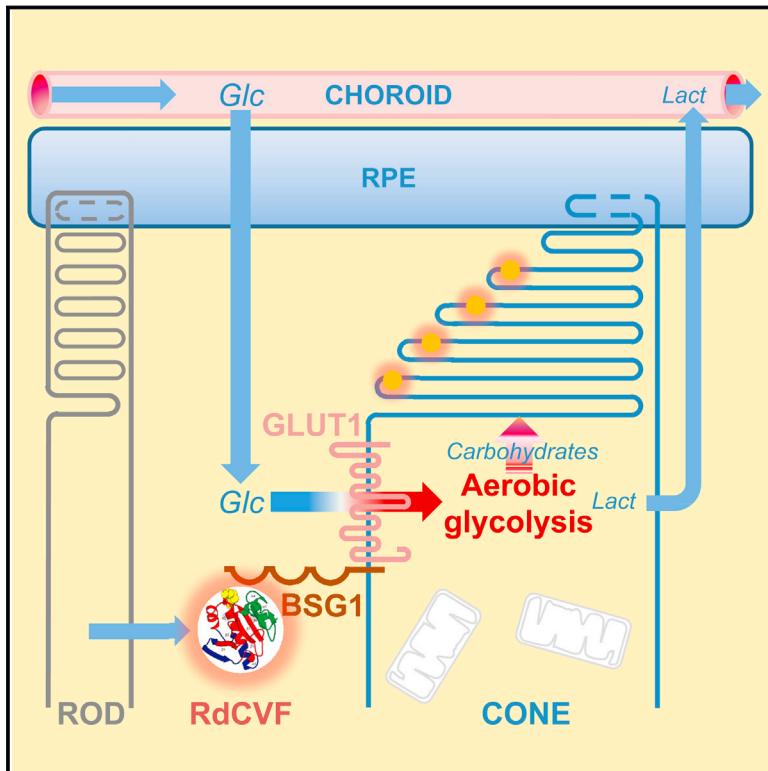
REFERENCES

- Arenkiel, B.R., Peca, J., Davison, I.G., Feliciano, C., Deisseroth, K., Augustine, G.J., Ehlers, M.D., and Feng, G. (2007). In vivo light-induced activation of neural circuitry in transgenic mice expressing channelrhodopsin-2. *Neuron* 54, 205–218.
- Bergles, D.E., Roberts, J.D., Somogyi, P., and Jahr, C.E. (2000). Glutamatergic synapses on oligodendrocyte precursor cells in the hippocampus. *Nature* 405, 187–191.
- Bergles, D.E., Jabs, R., and Steinhäuser, C. (2010). Neuron-glia synapses in the brain. *Brain Res. Brain Res. Rev.* 63, 130–137.
- Boyden, E.S., Zhang, F., Bamberg, E., Nagel, G., and Deisseroth, K. (2005). Millisecond-timescale, genetically targeted optical control of neural activity. *Nat. Neurosci.* 8, 1263–1268.
- Buckingham, S.C., Campbell, S.L., Haas, B.R., Montana, V., Robel, S., Ogunrinu, T., and Sontheimer, H. (2011). Glutamate release by primary brain tumors induces epileptic activity. *Nat. Med.* 17, 1269–1274.
- Campbell, S.L., Buckingham, S.C., and Sontheimer, H. (2012). Human glioma cells induce hyperexcitability in cortical networks. *Epilepsia* 53, 1360–1370.
- Charles, N.A., Holland, E.C., Gilbertson, R., Glass, R., and Kettenmann, H. (2011). The brain tumor microenvironment. *Glia* 59, 1169–1180.
- Chen, D.B., and Davis, J.S. (2003). Epidermal growth factor induces c-fos and c-jun mRNA via Raf-1/MEK1/ERK-dependent and -independent pathways in bovine luteal cells. *Mol. Cell. Endocrinol.* 200, 141–154.
- Creighton, C.J., Fu, X., Hennessy, B.T., Casa, A.J., Zhang, Y., Gonzalez-Angulo, A.M., Lluch, A., Gray, J.W., Brown, P.H., Hilsenbeck, S.G., et al. (2010). Proteomic and transcriptomic profiling reveals a link between the PI3K pathway and lower estrogen-receptor (ER) levels and activity in ER+ breast cancer. *Breast Cancer Res.* 12, R40.
- Cuddapah, V.A., Robel, S., Watkins, S., and Sontheimer, H. (2014). A neurocentric perspective on glioma invasion. *Nat. Rev. Neurosci.* 15, 455–465.
- Fan, Q.W., Cheng, C., Knight, Z.A., Haas-Kogan, D., Stokoe, D., James, C.D., McCormick, F., Shokat, K.M., and Weiss, W.A. (2009). EGFR signals to mTOR through PKC and independently of Akt in glioma. *Sci. Signal.* 2, ra4.
- Galvao, R.P., Kasina, A., McNeill, R.S., Harbin, J.E., Foreman, O., Verhaak, R.G., Nishiyama, A., Miller, C.R., and Zong, H. (2014). Transformation of quiescent adult oligodendrocyte precursor cells into malignant glioma through a multistep reactivation process. *Proc. Natl. Acad. Sci. USA* 111, E4214–E4223.
- Geha, S., Pallud, J., Junier, M.P., Devaux, B., Leonard, N., Chassoux, F., Chneiweiss, H., Daumas-Duport, C., and Varlet, P. (2010). NG2+/Olig2+ cells are the major cycle-related cell population of the adult human normal brain. *Brain Pathol.* 20, 399–411.
- Gerlee, P. (2013). The model muddle: in search of tumor growth laws. *Cancer Res.* 73, 2407–2411.
- Gibson, J.R., Huber, K.M., and Südhof, T.C. (2009). Neuroligin-2 deletion selectively decreases inhibitory synaptic transmission originating from fast-spiking but not from somatostatin-positive interneurons. *J. Neurosci.* 29, 13883–13897.
- Gibson, E.M., Purger, D., Mount, C.W., Goldstein, A.K., Lin, G.L., Wood, L.S., Inema, I., Miller, S.E., Bieri, G., Zuchero, J.B., et al. (2014). Neuronal activity promotes oligodendrogenesis and adaptive myelination in the mammalian brain. *Science* 344, 1252304.
- Gjorlund, M.D., Nielsen, J., Pankratova, S., Li, S., Korshunova, I., Bock, E., and Berezin, V. (2012). Neuroligin-1 induces neurite outgrowth through interaction with neuroligin-1 β and activation of fibroblast growth factor receptor-1. *FASEB J.* 26, 4174–4186.
- Gkogkas, C.G., Khoutorsky, A., Ran, I., Rampakakis, E., Nevarko, T., Weatherill, D.B., Vasuta, C., Yee, S., Truitt, M., Dallaire, P., et al. (2013). Autism-related deficits via dysregulated eIF4E-dependent translational control. *Nature* 493, 371–377.
- Glasgow, S.M., Zhu, W., Stolt, C.C., Huang, T.W., Chen, F., LoTurco, J.J., Neul, J.L., Wegner, M., Mohila, C., and Deneen, B. (2014). Mutual antagonism between Sox10 and NFIA regulates diversification of glial lineages and glioma subtypes. *Nat. Neurosci.* 17, 1322–1329.
- Gonzales, M., and Bowden, G.T. (2002). The role of PI 3-kinase in the UVB-induced expression of c-fos. *Oncogene* 21, 2721–2728.
- Greenberg, M.E., and Ziff, E.B. (1984). Stimulation of 3T3 cells induces transcription of the c-fos proto-oncogene. *Nature* 311, 433–438.
- Ichtchenko, K., Nguyen, T., and Südhof, T.C. (1996). Structures, alternative splicing, and neuroligin binding of multiple neuroligins. *J. Biol. Chem.* 271, 2676–2682.
- Ishiuchi, S., Tsuzuki, K., Yoshida, Y., Yamada, N., Hagimura, N., Okado, H., Miwa, A., Kurihara, H., Nakazato, Y., Tamura, M., et al. (2002). Blockage of Ca(2+)-permeable AMPA receptors suppresses migration and induces apoptosis in human glioblastoma cells. *Nat. Med.* 8, 971–978.
- Ishiuchi, S., Yoshida, Y., Sugawara, K., Aihara, M., Ohtani, T., Watanabe, T., Saito, N., Tsuzuki, K., Okado, H., Miwa, A., et al. (2007). Ca2+-permeable AMPA receptors regulate growth of human glioblastoma via Akt activation. *J. Neurosci.* 27, 7987–8001.
- Jamain, S., Quach, H., Betancur, C., Råstam, M., Colineaux, C., Gillberg, I.C., Soderstrom, H., Giros, B., Leboyer, M., Gillberg, C., and Bourgeron, T.; Paris Autism Research International Sibpair Study (2003). Mutations of the X-linked genes encoding neuroligins NLGN3 and NLGN4 are associated with autism. *Nat. Genet.* 34, 27–29.
- Johannessen, A.L., and Torp, S.H. (2006). The clinical value of Ki-67/MIB-1 labeling index in human astrocytomas. *Pathol. Oncol. Res.* 12, 143–147.
- Khuong-Quang, D.A., Buczkowicz, P., Rakopoulos, P., Liu, X.Y., Fontebasso, A.M., Bouffet, E., Bartels, U., Albrecht, S., Schwartzentruber, J., Letourneau, L., et al. (2012). K27M mutation in histone H3.3 defines clinically and biologically distinct subgroups of pediatric diffuse intrinsic pontine gliomas. *Acta Neuropathol.* 124, 439–447.
- Labrakakis, C., Patt, S., Hartmann, J., and Kettenmann, H. (1998). Functional GABA(A) receptors on human glioma cells. *Eur. J. Neurosci.* 10, 231–238.
- Lee, H.K., Xiang, C., Cazacu, S., Finniss, S., Kazimirsky, G., Lemke, N., Lehman, N.L., Rempel, S.A., Mikkelsen, T., and Brodie, C. (2008). GRP78 is overexpressed in glioblastomas and regulates glioma cell growth and apoptosis. *Neuro-oncol.* 10, 236–243.
- Liu, C., Sage, J.C., Miller, M.R., Verhaak, R.G., Hippenmeyer, S., Vogel, H., Foreman, O., Bronson, R.T., Nishiyama, A., Luo, L., and Zong, H. (2011). Mosaic analysis with double markers reveals tumor cell of origin in glioma. *Cell* 146, 209–221.
- Magnon, C., Hall, S.J., Lin, J., Xue, X., Gerber, L., Freedland, S.J., and Frenette, P.S. (2013). Autonomic nerve development contributes to prostate cancer progression. *Science* 341, 1236361.
- Mehta, S., Huillard, E., Kesari, S., Maire, C.L., Golebiowski, D., Harrington, E.P., Alberta, J.A., Kane, M.F., Theisen, M., Ligon, K.L., et al. (2011). The central nervous system-restricted transcription factor Olig2 opposes p53

- responses to genotoxic damage in neural progenitors and malignant glioma. *Cancer Cell* 19, 359–371.
- Monje, M., Mitra, S.S., Freret, M.E., Raveh, T.B., Kim, J., Masek, M., Attena, J.L., Li, G., Haddix, T., Edwards, M.S., et al. (2011). Hedgehog-responsive candidate cell of origin for diffuse intrinsic pontine glioma. *Proc. Natl. Acad. Sci. USA* 108, 4453–4458.
- Patel, A.P., Tirosh, I., Trombetta, J.J., Shalek, A.K., Gillespie, S.M., Wakimoto, H., Cahill, D.P., Nahed, B.V., Curry, W.T., Martuza, R.L., et al. (2014). Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science* 344, 1396–1401.
- Peixoto, R.T., Kunz, P.A., Kwon, H., Mabb, A.M., Sabatini, B.L., Philpot, B.D., and Ehlers, M.D. (2012). Transsynaptic signaling by activity-dependent cleavage of neuroligin-1. *Neuron* 76, 396–409.
- Persson, A.I., Petritsch, C., Swartling, F.J., Itsara, M., Sim, F.J., Auvergne, R., Goldenberg, D.D., Vandenberg, S.R., Nguyen, K.N., Yakovenko, S., et al. (2010). Non-stem cell origin for oligodendroglioma. *Cancer Cell* 18, 669–682.
- Pyonteck, S.M., Akkari, L., Schuhmacher, A.J., Bowman, R.L., Sevenich, L., Quail, D.F., Olson, O.C., Quick, M.L., Huse, J.T., Teijeiro, V., et al. (2013). CSF-1R inhibition alters macrophage polarization and blocks glioma progression. *Nat. Med.* 19, 1264–1272.
- Rothstein, J.D., Dykes-Hoberg, M., Pardo, C.A., Bristol, L.A., Jin, L., Kuncl, R.W., Kanai, Y., Hediger, M.A., Wang, Y., Schielke, J.P., and Welty, D.F. (1996). Knockout of glutamate transporters reveals a major role for astroglial transport in excitotoxicity and clearance of glutamate. *Neuron* 16, 675–686.
- Rothwell, P.E., Fuccillo, M.V., Maxeiner, S., Hayton, S.J., Gokce, O., Lim, B.K., Fowler, S.C., Malenka, R.C., and Südhof, T.C. (2014). Autism-associated neuroligin-3 mutations commonly impair striatal circuits to boost repetitive behaviors. *Cell* 158, 198–212.
- Samarelli, A.V., Riccitelli, E., Bizzozero, L., Silveira, T.N., Seano, G., Pergolizzi, M., Vitagliano, G., Cascone, I., Carpentier, G., Bottos, A., et al. (2014). Neuroligin 1 induces blood vessel maturation by cooperating with the $\alpha 6$ integrin. *J. Biol. Chem.* 289, 19466–19476.
- Scherer, H.J. (1938). Structural development in gliomas. *Am. J. Cancer* 34, 333–351.
- Schwartzentruber, J., Korshunov, A., Liu, X.Y., Jones, D.T., Pfaff, E., Jacob, K., Sturm, D., Fontebasso, A.M., Quang, D.A., Tönjes, M., et al. (2012). Driver mutations in histone H3.3 and chromatin remodelling genes in paediatric glioblastoma. *Nature* 482, 226–231.
- Seifert, S., and Sontheimer, H. (2014). Bradykinin enhances invasion of malignant glioma into the brain parenchyma by inducing cells to undergo amoeboid migration. *J. Physiol.* 592, 5109–5127.
- Silver, D.J., Siebzehnrubl, F.A., Schildts, M.J., Yachnis, A.T., Smith, G.M., Smith, A.A., Scheffler, B., Reynolds, B.A., Silver, J., and Steindler, D.A. (2013). Chondroitin sulfate proteoglycans potently inhibit invasion and serve as a central organizer of the brain tumor microenvironment. *J. Neurosci.* 33, 15603–15617.
- Stopczynski, R.E., Normolle, D.P., Hartman, D.J., Ying, H., DeBerry, J.J., Bielefeldt, K., Rhim, A.D., DePinho, R.A., Albers, K.M., and Davis, B.M. (2014). Neuroplastic changes occur early in the development of pancreatic ductal adenocarcinoma. *Cancer Res.* 74, 1718–1727.
- Sturm, D., Witt, H., Hovestadt, V., Khuong-Quang, D.A., Jones, D.T., Konermann, C., Pfaff, E., Tönjes, M., Sill, M., Bender, S., et al. (2012). Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. *Cancer Cell* 22, 425–437.
- Südhof, T.C. (2008). Neuroligins and neuexins link synaptic function to cognitive disease. *Nature* 455, 903–911.
- Sugiarto, S., Persson, A.I., Munoz, E.G., Waldhuber, M., Lamagna, C., Andor, N., Hanecker, P., Ayers-Ringler, J., Phillips, J., Siu, J., et al. (2011). Asymmetry-defective oligodendrocyte progenitors are glioma precursors. *Cancer Cell* 20, 328–340.
- Suzuki, K., Hayashi, Y., Nakahara, S., Kumazaki, H., Prox, J., Horiuchi, K., Zeng, M., Tanimura, S., Nishiyama, Y., Osawa, S., et al. (2012). Activity-dependent proteolytic cleavage of neuroligin-1. *Neuron* 76, 410–422.
- Synowitz, M., Ahmann, P., Matyash, M., Kuhn, S.A., Hofmann, B., Zimmer, C., Kirchhoff, F., Kiwit, J.C., and Kettenmann, H. (2001). GABA(A)-receptor expression in glioma cells is triggered by contact with neuronal cells. *Eur. J. Neurosci.* 14, 1294–1302.
- Tabuchi, K., Blundell, J., Etherton, M.R., Hammer, R.E., Liu, X., Powell, C.M., and Südhof, T.C. (2007). A neuroligin-3 mutation implicated in autism increases inhibitory synaptic transmission in mice. *Science* 318, 71–76.
- Ullrich, B., Ushkaryov, Y.A., and Südhof, T.C. (1995). Cartography of neuexins: more than 1000 isoforms generated by alternative splicing and expressed in distinct subsets of neurons. *Neuron* 14, 497–507.
- Varoqueaux, F., Aramuni, G., Rawson, R.L., Mohrmann, R., Missler, M., Gottmann, K., Zhang, W., Südhof, T.C., and Brose, N. (2006). Neuroligins determine synapse maturation and function. *Neuron* 51, 741–754.
- Verhaak, R.G., Hoadley, K.A., Purdom, E., Wang, V., Qi, Y., Wilkerson, M.D., Miller, C.R., Ding, L., Golub, T., Mesirov, J.P., et al.; Cancer Genome Atlas Research Network (2010). Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* 17, 98–110.
- Wang, H., Peca, J., Matsuzaki, M., Matsuzaki, K., Noguchi, J., Qiu, L., Wang, D., Zhang, F., Boyden, E., Deisseroth, K., et al. (2007). High-speed mapping of synaptic connectivity using photostimulation in Channelrhodopsin-2 transgenic mice. *Proc. Natl. Acad. Sci. USA* 104, 8143–8148.
- Wang, Y., Yang, J., Zheng, H., Tomasek, G.J., Zhang, P., McKeever, P.E., Lee, E.Y., and Zhu, Y. (2009). Expression of mutant p53 proteins implicates a lineage relationship between neural stem cells and malignant astrocytic glioma in a murine model. *Cancer Cell* 15, 514–526.
- Wu, G., Broniscer, A., McEachron, T.A., Lu, C., Paugh, B.S., Becksfort, J., Qu, C., Ding, L., Huether, R., Parker, M., et al.; St. Jude Children's Research Hospital–Washington University Pediatric Cancer Genome Project (2012). Somatic histone H3 alterations in pediatric diffuse intrinsic pontine gliomas and non-brainstem glioblastomas. *Nat. Genet.* 44, 251–253.
- Yizhar, O., Fenno, L.E., Davidson, T.J., Mogri, M., and Deisseroth, K. (2011). Optogenetics in neural systems. *Neuron* 71, 9–34.
- Zhang, Y., Chen, K., Sloan, S.A., Bennett, M.L., Scholze, A.R., O'Keefe, S., Phatnani, H.P., Guarnieri, P., Caneda, C., Ruderisch, N., et al. (2014). An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. *J. Neurosci.* 34, 11929–11947.
- Zhao, C.M., Hayakawa, Y., Kodama, Y., Muthupalani, S., Westphalen, C.B., Anderson, G.T., Flatberg, A., Johannessen, H., Friedman, R.A., Renz, B.W., et al. (2014). Denervation suppresses gastric tumorigenesis. *Sci. Transl. Med.* 6, 250ra115.

Rod-Derived Cone Viability Factor Promotes Cone Survival by Stimulating Aerobic Glycolysis

Graphical Abstract



Authors

Najate Aït-Ali, Ram Fridlich, ..., José-Alain Sahel, Thierry Lévillard

Correspondence

thierry.leveillard@inserm.fr

In Brief

The rod-derived cone viability factor RdCVF promotes retinal cone survival by accelerating the entry of glucose into photoreceptors and enhancing aerobic glycolysis. RdCVF acts by binding to the cell-surface complex BSG1/GLUT1, a pathway also used by fast dividing cancer cells.

Highlights

- Basigin-1 is the cell-surface receptor for RdCVF
- RdCVF stimulates glucose uptake by cones
- RdCVF-mediated glucose uptake stimulates aerobic glycolysis and cone survival

Accession Numbers

GSE62020
PXD001384
PXD001715



Rod-Derived Cone Viability Factor Promotes Cone Survival by Stimulating Aerobic Glycolysis

Najate Aït-Ali,^{1,2,3,13} Ram Fridlich,^{1,2,3,13} Géraldine Millet-Puel,^{1,2,3,13} Emmanuelle Clérin,^{1,2,3,13} François Delalande,^{4,5} Céline Jaillard,^{1,2,3} Frédéric Blond,^{1,2,3} Ludivine Perrocheau,^{1,2,3} Sacha Reichman,^{1,2,3} Leah C. Byrne,⁶ Anne Olivier-Bandini,⁷ Jacques Bellalou,⁸ Emmanuel Moyse,⁹ Frédéric Bouillaud,^{10,11,12} Xavier Nicol,^{1,2,3} Deniz Dalkara,^{1,2,3} Alain van Dorsselaer,^{4,5} José-Alain Sahel,^{1,2,3} and Thierry Léveillard^{1,2,3,*}

¹INSERM, U968, 75012 Paris, France

²Sorbonne Universités, UPMC Univ Paris 06, UMR_S 968, Institut de la Vision, 75012 Paris, France

³CNRS, UMR_7210, 75012 Paris, France

⁴BioOrganic Mass Spectrometry Laboratory (LSMBO), IPHC, Université de Strasbourg, 25 rue Becquerel, 67087 Strasbourg, France

⁵IPHC, CNRS, UMR7178, 67087 Strasbourg, France

⁶Helen Wills Neuroscience Institute, University of California Berkeley, Berkeley, CA 94720, USA

⁷Sanofi R&D, 1 Avenue Pierre Brossolette, 91385 Chilly-Mazarin, France

⁸Institut Pasteur, Platform 5 Production of Recombinant Proteins and Antibodies, 75724 Paris Cedex 15, France

⁹Unité de Physiologie de la Reproduction et des Comportements (PRC), UMR-85 INRA, Centre INRA de Tours, Université François Rabelais de Tours, 37380 Nouzilly, France

¹⁰Inserm, U1016, Institut Cochin, 75014 Paris, France

¹¹Cnrs, UMR8104, 75014 Paris, France

¹²Université Paris Descartes, Sorbonne Paris Cité, 75014 Paris, France

¹³Co-first author

*Correspondence: thierry.leveillard@inserm.fr

<http://dx.doi.org/10.1016/j.cell.2015.03.023>

SUMMARY

Rod-derived cone viability factor (RdCVF) is an inactive thioredoxin secreted by rod photoreceptors that protects cones from degeneration. Because the secondary loss of cones in retinitis pigmentosa (RP) leads to blindness, the administration of RdCVF is a promising therapy for this untreatable neurodegenerative disease. Here, we investigated the mechanism underlying the protective role of RdCVF in RP. We show that RdCVF acts through binding to Basigin-1 (BSG1), a transmembrane protein expressed specifically by photoreceptors. BSG1 binds to the glucose transporter GLUT1, resulting in increased glucose entry into cones. Increased glucose promotes cone survival by stimulation of aerobic glycolysis. Moreover, a missense mutation of RdCVF results in its inability to bind to BSG1, stimulate glucose uptake, and prevent secondary cone death in a model of RP. Our data uncover an entirely novel mechanism of neuroprotection through the stimulation of glucose metabolism.

INTRODUCTION

RdCVF, a truncated thioredoxin-like protein lacking thioredoxin reductase activity, was identified by high content screening of a mouse retinal cDNA library on cone-enriched cultures from chicken embryos (Léveillard et al., 2004). RdCVF is an alternative splice variant of the nucleoredoxin-like 1 (*Nxn1*) gene, whose

other splice product is RdCVFL, an active thioredoxin that protects its binding partner, the microtubule associated protein TAU, from oxidation and aggregation (Elachouri et al., 2015; Fridlich et al., 2009). *Nxn1*^{−/−} mice experience an age-dependent loss of rod and cone function and cone degeneration. Rods and cones of *Nxn1*^{−/−} mice are also hypersensitive to oxidative stress (Cronin et al., 2010). The expression of *Nxn1* is rod dependent in the retina and is severely reduced after rod death in retinitis pigmentosa (RP) (Delyfer et al., 2011; Reichman et al., 2010).

We have demonstrated that RdCVF, but not RdCVFL, protects cone function in several genetically distinct models of RP, targeting the most debilitating step in that untreatable disease (Byrne et al., 2015; Léveillard et al., 2004; Yang et al., 2009). In patients suffering from RP, the most common form of inherited retinal degeneration, vision loss develops in two successive steps. Early in adult life, these patients lose the ability to see in dim light conditions (night vision loss), corresponding to the loss of function and degeneration of rods. This is felt as a minor handicap, especially in individuals affected by congenital stationary night blindness, an inherited retinal disease characterized exclusively by lack of rod function. In well-illuminated environments, these people retain an almost normal way of life (Dryja et al., 1996). For RP patients, the disease then progresses through another debilitating step resulting from the loss of function and degeneration of cones that dominate at the center of the retina and represent 5% of all photoreceptors in human and most mammals. Treating RP patients by replacing the expression of RdCVF will not correct the causative gene defect but should maintain cone-mediated central vision, potentially benefiting an estimated 1.5 million people worldwide (Wright, 1997).

Thioredoxins catalyze the reduction of disulfide bonds in many proteins (Holmgren, 1985). Human thioredoxin-1, originally

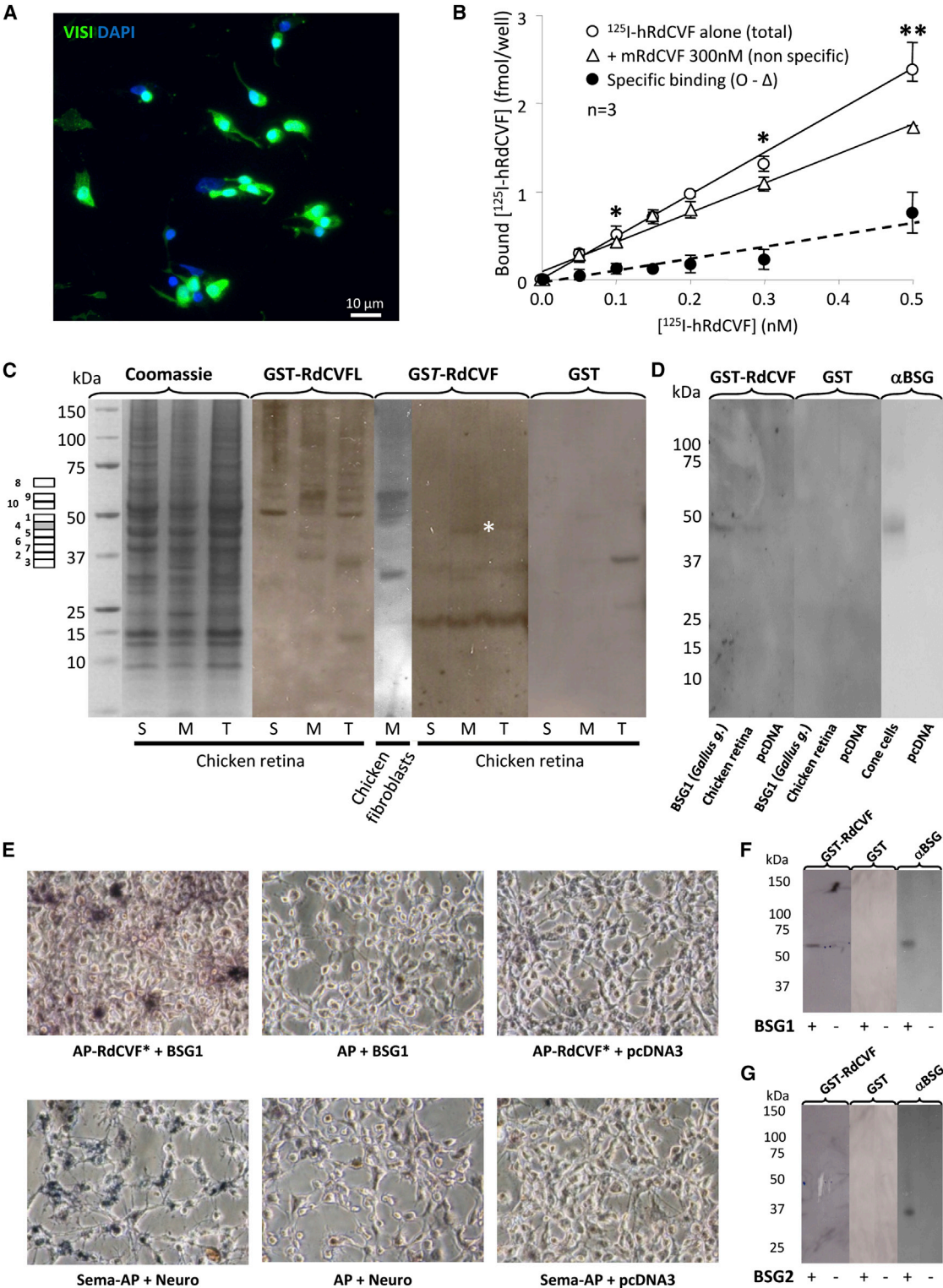


Figure 1. Rod-Derived Cone Viability Factor Binds to Basigin-1
(A) Immunocytochemical analysis of the cone-enriched cultures with anti-visinin (VISI) antibodies.
(B) Binding of ¹²⁵I-labeled human RdCVF (hRdCVF) to cone-enriched culture cells and its competitive inhibition by unlabelled mouse recombinant RdCVF (mRdCVF). Open symbols correspond to cell-bound radioactivity measurements after incubation in the absence (“total” binding) or in the presence (non-specific binding) of mRdCVF. Asterisks indicate significant differences between total and non-specific binding (p < 0.05).

(legend continued on next page)

identified as the secreted protein adult T cell leukemia-derived factor (ADF), has been implicated in a wide variety of redox regulations in both intracellular and extracellular compartments (Matsuo and Yodoi, 2013). Thioredoxins are secreted by an unknown leader-less pathway (Rubartelli et al., 1992). Extracellular thioredoxins, including the enzymatically inactive truncated of thioredoxin-1 (TRX80) exert paracrine effects (Pekkari et al., 2003). Nevertheless, receptors for extracellular thioredoxins are scarce in the literature. The tumor necrosis factor receptor TNFRSF8 is the principal target of thioredoxin-1 on lymphocytes (Schwertassek et al., 2007), and the TRPC5 channel is activated by extracellular thioredoxin-1 (Xu et al., 2008). The receptor for TRX80 is presently unknown. Given the paucity of data on extracellular thioredoxin signaling pathways, we used a far-western blotting approach to identify basigin-1 as the transducing RdCVF receptor on the surface of cones. We then revealed that RdCVF interacts with a complex formed by basigin-1 and the glucose transporter GLUT1 to stimulate aerobic glycolysis and induce cone survival.

RESULTS

Basigin-1 Is the Cell-Surface Receptor for RdCVF

Cone-enriched cultures from chicken embryos are composed of ~80% cone photoreceptors as seen through the labeling of visinin, a chicken photoreceptor marker (Figure 1A). Biologically active synthetic human RdCVF protein (Yang et al., 2009) was labeled with ¹²⁵Iodine and incubated with cone-enriched cultures in the presence or the absence of excess (300 nM) non-radioactive mouse synthetic RdCVF protein. After washing, cell-bound radioactive material was isolated by filtration and counted. Specific receptor binding was evidenced as the part of cell-bound radioligand that is inhibited by competition with non-radioactive RdCVF (Figure 1B). No specific binding was observed in primary retinal pigmented epithelial cells or COS-1 cells (Figures S1A and S1B). We then used a far-western blotting approach to identify proteins on the surface of the cones that bind to RdCVF. Soluble, membrane-bound and total fractions from chicken retina were run on a gel, transferred to nitrocellulose membranes, and incubated with GST, GST-RdCVF, or GST-RdCVFL protein. Chicken embryonic fibroblast cultures were used as a negative control. Binding was then revealed using anti-GST antibodies (Figure 1C). A stronger specific signal was observed in the membrane fraction (M) of chicken retina with GST-RdCVF compared to GST-RdCVFL. We sliced the

membrane fraction lane from a Coomassie-stained gel into ten pieces for mass spectrometry/mass spectrometry (MS/MS) analysis. Slices 4 and 5, aligning with the candidate signal, contain 30 major polypeptides, among which only two are transmembrane proteins (Table S1A; Figure S1C). When we repeated the experiment using more recent instruments, we identified 26 integral component of membrane proteins among them basigin-1 and the other candidate, ATP1B3 in similar gel slides prepared from cone-enriched cultures (Table S1B). To validate the interaction of RdCVF with basigin-1 (previously known as basigin-2; Ochrietor et al., 2003), COS-1 cells were transfected with chicken basigin-1 cDNA or a negative control (Figure 1D). Far-western blotting on membrane fractions from these samples, following incubation with GST-RdCVF, produced a signal matching that of basigin-1 as revealed by western blotting on membrane fraction of cone-enriched cultures.

Using an alternative strategy, we demonstrated that the RdCVF-BSG1 interaction takes place in cellular context. A RdCVF-alkaline phosphatase (AP) fusion protein was produced by transfection of HEK293 cells (Figures S1E and S1F), and the resulting RdCVF-containing conditioned media were incubated with COS-1 cells previously transfected with basigin-1. AP staining, indicating binding of RdCVF-AP, was observed only when COS-1 cells express basigin-1 (Figure 1E). The interaction of semaphorin with its receptor neuropilin was used as a positive control. No binding was observed with the second candidate, ATP1B3 (Figure S1D).

The basigin gene encodes for two products by alternative splicing. Basigin-2, a protein with two extracellular immunoglobulin domains, is widely expressed, while basigin-1, a protein with a third immunoglobulin domain (IgD), is expressed specifically in the retina (Ochrietor et al., 2003). Using far-western blotting and AP fusion protein assay, we found that RdCVF interacts with basigin-1, but not basigin-2 (Figures 1F, 1G, and S1D). The positive signal matches that of basigin-1 revealed by western blotting.

We also explored the effect of silencing basigin expression on RdCVF-mediated cone-enriched cultures survival. Immunocytochemical labeling using a monoclonal antibody that recognizes basigin-1 and basigin-2 revealed basigin expression by cone-enriched cultures at the cell surface (Figure 2A). We then measured the protective effect of RdCVF in cone-enriched cultures after small interfering RNA (siRNA) silencing of basigin expression (basigin-1 + 2). In this system, the post-mitotic primary cells degenerate over a period of 7 days (Figures S2A–S2C). siRNA was validated using luciferase reporter assay

binding) of micromolar non-radioactive RdCVF. Black symbol is the specific radioligand binding, as calculated by difference between total and non-specific measurements. ANOVA, SDs. n = 3.

(C) Far-western blotting analysis of fractions from chicken retina and embryonic fibroblast culture using GST, GST-RdCVF, and GST-RdCVFL. The candidate band is indicated with an asterisk. The numbers on the left correspond to slices of the gel that were excised for MS/MS analysis. The candidate band is located in gel slice 4 in gray. S, M, and T, soluble, membrane, and total fraction, respectively.

(D) Far-western blotting of membrane extracts from COS-1 cells transfected with chicken basigin-1 or negative control (pcDNA3). Western blotting analysis of the expression of basigin-1 (chicken) by cone-enriched culture cells.

(E) Detection of the interaction between RdCVF and basigin-1 in COS-1-transfected cells using alkaline phosphatase fusion proteins (AP). The interaction of semaphorin (Sema) with its receptor neuropilin (Neuro) was used as positive control.

(F) Far-western blotting of membrane extracts of COS-1 cells transfected with mouse basigin-1 or pcDNA3. Western blotting analysis with anti-BSG (mouse) antibodies.

(G) Far-western blotting of membrane extracts from COS-1 cells transfected with mouse basigin-2 or pcDNA3. Western blotting analysis with anti-BSG (mouse). See also Figure S1 and Tables S1A and S1B.

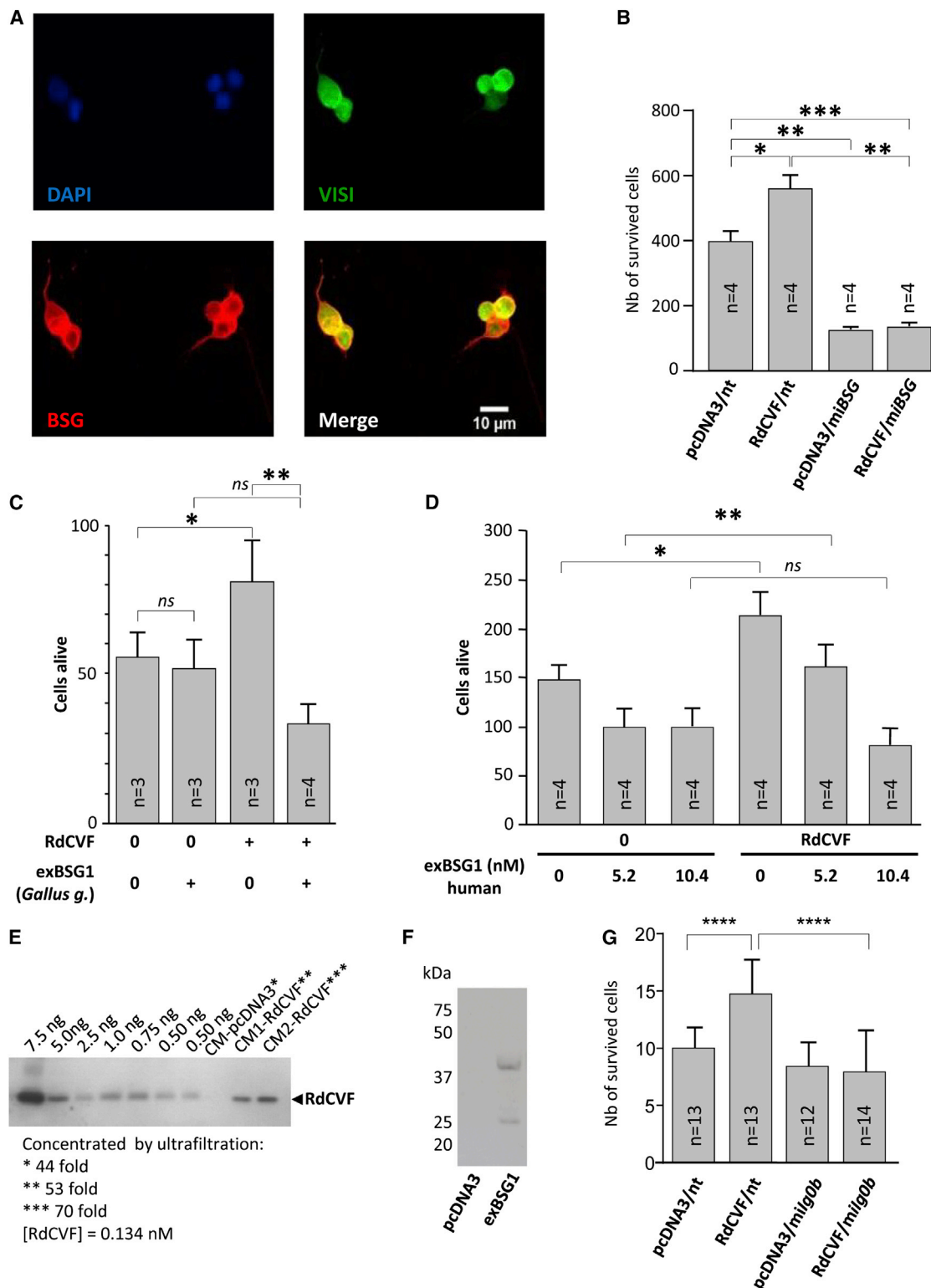


Figure 2. Basigin Is Involved in RdCVF-Mediated Cone Survival

(A) Immunocytochemical analysis of cone-enriched cultures with anti-basigin and anti-visinin antibodies.

(B) Effect of silencing basigin in cone-enriched cultures on RdCVF-mediated cell survival. nt, non-targeting siRNA. Tukey's test, SDs. n = 4.

(C) Competitive effect of the extracellular domain of chicken basigin-1 (exBSG1) in cone-enriched cultures on RdCVF-mediated cell survival. Fisher's test, SD, n = 3/4.

(legend continued on next page)

(Figure S2D). The number of transduced cone-enriched culture cells was quantified using a RFP reporter (Extended Experimental Procedures). Silencing of basigin reduced cell survival mediated by ectopic RdCVF, while a non-targeting (nt) siRNA construct did not (Figures 2B and S2E). The reduction of cell survival for unstimulated cone-enriched cultures suggests that an RdCVF-independent pathway may also be involved. To further explore that possibility, we used a competition assay. Cone-enriched cultures were incubated with conditioned media (CM) from COS-1 cells transfected with RdCVF and the extracellular domain of basigin-1 (exBSG1) from chicken. Co-expression of RdCVF and exBSG1 abolished cell survival, while, in contrast, expression of exBSG1 alone had no impact on cone-enriched cultures cell survival (Figure 2C). We also showed that purified human exBSG1 competes for RdCVF-mediated survival in a dose-dependent manner (Figure 2D). A semiquantitative western blotting approach revealed that the concentration of RdCVF in the conditioned medium is 0.134 nM (Figure 2E). This corresponds to a 39- and 78-fold excess over the recombinant exBSG1. We also detected exBSG1 (chicken) in the conditioned medium of COS-1-transfected cells (Figure 2F). We confirmed that RdCVF-mediated cone survival depends of basigin-1 (Figure 2G).

RdCVF Stimulates Cone Survival by Increasing Glucose Uptake

Basigin-1 possesses a single transmembrane domain and a short cytoplasmic domain of 40–44 residues unlinked to any known signaling pathway. We co-immunoprecipitated basigin-1-interacting proteins from chicken retina and identified them by MS/MS (Figures 3A and S3A; Table S1C). To analyze the data, we subtracted the proteins also identified in the negative control (immunoglobulin G [IgG]). Among the five identified proteins, apart from basigin-1 itself, we focused on the glucose transporter GLUT1 (SLC2A1). GLUT1 antibodies co-immunoprecipitated basigin (BSG1 and BSG2) from membrane fraction of chicken retina (Figure 3B). We validated the interaction between basigin-1 and GLUT1 using fluorescence resonance energy transfer (FRET) (Figure 3C). FRET signal was detected in cells transfected with GLUT1-CFP and BSG1-EYFP fusion proteins revealing their close vicinity and interaction. Interestingly, basigin-1 also interacted with the lactate transporter MCT1 (SLC16A1) as previously reported for basigin-2 (Kirk et al., 2000). The results were significant for basigin-1 interaction with both GLUT1 and MCT1 (Figure 3D). The analysis of basigin-1 sub-cellular localization in transfected-HEK293 cells indicated that the interaction was not of the same nature. Basigin-1 co-localizes in the cells with MCT1 when co-expressed, while the expression pattern of basigin-1 is mainly on cell surface and reticulated when co-expressed with GLUT1 (Figures S3B and S3C). We also observed that GLUT1, in addition to basigin-1, was expressed in cone-enriched cultures (Figure 3E).

Cone survival is mediated by RdCVF and not by the thiodoxin RdCVFL, the other product of the *Nxn1* gene (Byrne et al., 2015). Using cone-enriched culture system, we observed that RdCVF increases cone survival but not RdCVFL, even if a trend was observed (Figure 4A). Using a fluorescent non-metabolized analog of glucose, 2-NBDG, we measured the entry of glucose into cells (Chen et al., 2015). The fluorescence of 2-NBDG was detected in the cytoplasm of the cell after 10 min incubation (Figure 4B). Conditioned medium containing RdCVF increased the uptake of glucose in cone-enriched culture cells, while RdCVFL did not (Figure 4C). The uptake of 2-NBDG is linear between 2.5 and 12.5 min (Figure 4D). Glucose uptake by cone-enriched culture cells was reduced to its unstimulated level when basigin was silenced (Figures 4E and 4F). Similar results were obtained by silencing basigin-1 through the use of two distinct siRNAs targeting the specific exon of basigin-1 encoding Ig0 (Figure 4G). RdCVF-mediated glucose uptake depended on the expression of GLUT1 as demonstrated by the effect of co-expressing two distinct siRNA targeting GLUT1 (Figure 4H), which both reduced the uptake of glucose by cone cells to unstimulated level. The fact that the interaction of basigin-1 and GLUT1 takes place at the cell surface suggested that RdCVF acts directly on that complex. However, the addition of RdCVF to cone-enriched culture cells did not modify the concentration of GLUT1 at the cell surface as shown by exposure to a non-permeable crosslinking reagent was used to purify cell-surface proteins (Figure 4I).

We obtained additional evidence of RdCVF/BSG1/GLUT1 complex formation at the surface of the cell by studying glucose uptake kinetics in cone-enriched culture cells after depletion of RdCVF protein. In this assay, conditioned medium containing RdCVF is removed as well as glucose before 2-NBDG is added. One hour after depletion, glucose uptake was higher in the cells cultured in the presence of RdCVF-containing conditioned medium than in the negative control, pcDNA3 (Figure 4J). Interestingly, the amplitude of the effect was reduced 2 hr after depletion and disappeared after 3 hr. The reduction was not due to a difference in cell survival as seen by the level of expression of visinin in both conditions when cells were cultured for 4 days (Figure 4K). The total level of expression of GLUT1 by cone-enriched culture cells was not modulated in these conditions.

The *Bsg*^{−/−} mouse is not the appropriate model to study RdCVF signaling in vivo. The deletion of the basigin gene in the mouse causes many defects, including the loss of photoreceptors (Hori et al., 2000), but the effect on photoreceptors might be first attributed to a defect in lactate transport by retinal pigmented epithelial cells (Daniele et al., 2008). In order to circumvent this problem, we studied the properties of an RdCVF variant, corresponding to a missense mutation found in a Leber congenital amaurosis patient (Hanein et al., 2006). In this mutant, an E to K mutation at position 64 is located in a potential interface in the structural model of RdCVF (Chalmel et al., 2007). The

(D) Competitive effect of the purified extracellular domain of human basigin-1 (exBSG1) in cone-enriched cultures on RdCVF-mediated cell survival. See also Figure S2. Tukey's test, SD, n = 4.

(E) Semiquantitative estimation of the concentration of RdCVF in the conditioned medium of COS-1-transfected cells. CM1 and CM2, two independent experiments.

(F) Western blotting analysis of the conditioned medium of COS-1 cells transfected with pexBSG1 (chicken).

(G) Effect of silencing specifically basigin-1 in cone-enriched cultures on RdCVF-mediated cell survival. nt, non-targeting siRNA. Dunnett's test, SD, n = 12–14.

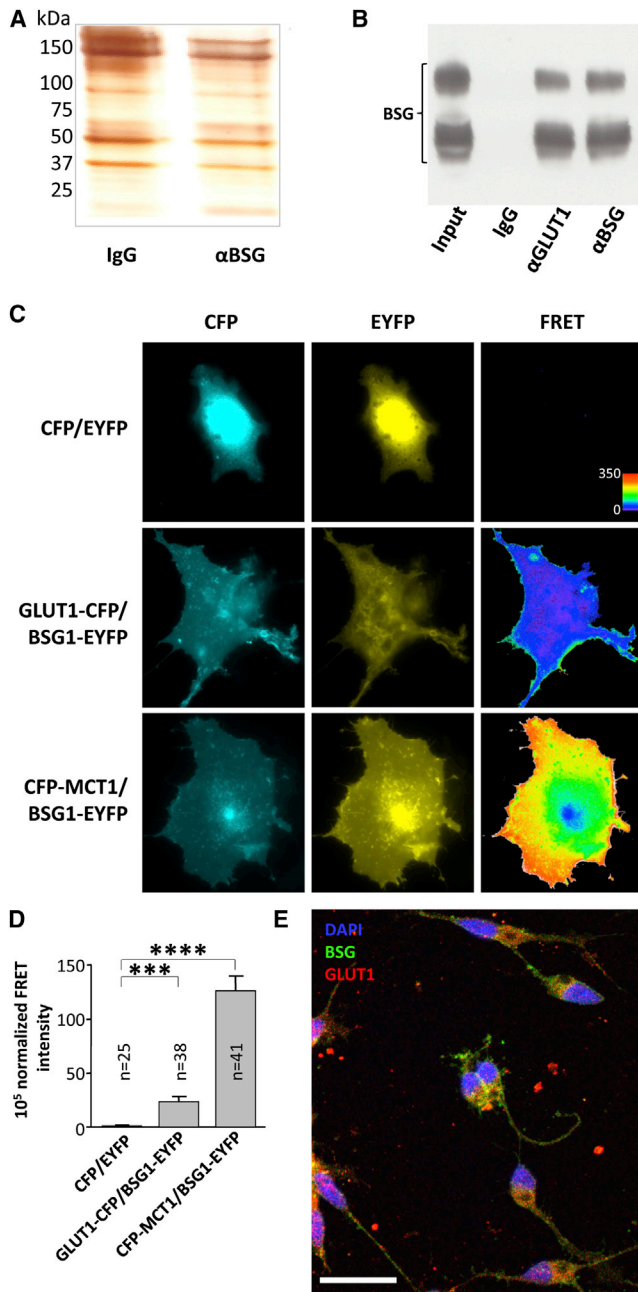


Figure 3. Basigin-1 Interacts with the Glucose Transporter GLUT1
 (A) Silver-stained gel of membrane proteins from chicken retina co-immunoprecipitated with anti-basigin antibody.
 (B) Co-immunoprecipitation of chicken basigin with anti-GLUT1 antibodies. IgG, negative control.
 (C) Interaction of basigin-1 with GLUT1 and MCT1 using FRET. Low to high normalized FRET intensity is color coded from black (no FRET) to red. Lower but significant FRET is coded in violet-blue.
 (D) Quantification of the intensity of the FRET signal. Dunnett's test, SEM, $n = 25/38/41$.
 (E) Immunocytochemical analysis of cone-enriched cultures with anti-BSG and anti-GLUT1 antibodies.
 Scale bar, 10 μ m. See also Figure S3 and Table S1C.

NXNL1 gene also encodes for two polypeptides, RdCVFL and RdCVF, by alternative splicing (Figures S4A–S4C).

Normal human RdCVF wild-type (hRdCVF wt) protected cone-enriched culture cells similarly to mouse RdCVF (mRdCVF), whereas the E64K variant (hRdCVF mut) was defective in this activity (Figure 5A). This lack of activity was not the result of a deficit in expression or secretion of RdCVF mut (Figure 5B). However, RdCVF mut was not able to stimulate glucose uptake in cone-enriched culture, in contrast to RdCVF wt (Figure 5C). Moreover, when used as a probe, GST-RdCVF mut (human) did not interact, or very weakly, with purified exBSG1 by far-western blotting (Figure 5D). Purified TAU was used as negative control (Fridlich et al., 2009). This did not reflect differences in production or purification of the mutant versus the normal recombinant protein (Figure 5E) but rather indicated a decrease in affinity of the mutant RdCVF protein for basigin-1.

We used this missense mutant to explore the role of RdCVF interaction with basigin-1 in vivo. The human cDNAs of RdCVF wt and mut were cloned into a construct engineered to produce GFP and RdCVF via a self-cleaving 2A peptide (GFP-2A-RdCVF) and containing AAV packaging elements. COS-1 cells were transfected with GFP-2A-RdCVF and a GFP negative control. Conditioned medium was then analyzed by western blotting (Figure 5F). Anti-GFP antibody identified a 26 kDa protein (GFP) in all lanes and a 40 kDa band corresponding to uncleaved GFP-2A-RdCVF wt and GFP-2A-RdCVF mut, while anti-RdCVF recognized the same 40 kDa band and a 12 kDa protein, RdCVF, which was present in similar amount for wild-type and mutant plasmids. The conditioned medium from transfected COS-1 cells was then applied to cone-enriched cultures. As compared to culture medium alone (0), GFP was slightly toxic (Figure 5G). RdCVF wt doubled the number of living cone-enriched culture cells, while RdCVF mut had no effect.

The constructs were then packaged in AAV9-2YF virus, which was injected via intracardiac injection in *rd1* mice, a well-studied model of rod-cone degeneration (Léveillard et al., 2004). Intravascular administration of AAV9 has been previously used to deliver RdCVF to the retina (Byrne et al., 2015). In *rd1* mice, a rod-specific recessive *Pde6b* mutation triggers rod degeneration by post-natal day 21 (PN21), which is followed by the non-cell autonomous degeneration of cones. Injections were made in *rd1* mice at PN4, and by PN38, the ocular fundus of the treated mice revealed transgene expression through GFP fluorescence (Figure 5H). Animals were sacrificed at PN49, cones were labeled with peanut agglutinin (PNA), and the cone density was measured using an automated platform (Clérin et al., 2011). Cone density across the *rd1* retina was similar for animals treated with AAV-GFP and AAV-GFP-2A-RdCVF mut, while the density of the cones in mice treated with AAV-GFP-2A-RdCVF WT was significantly higher (Figures 5I, S4D, and S4E). We then measured the expression of RdCVF by RT-PCR in contralateral eyes with primers that do not discriminate RdCVF wt and mut cDNAs. The expression of RdCVF mRNA was found to be equivalent for mice treated with AAV-GFP-2A-RdCVF wt and mut (Figure 5J). In the P23H rat, a dominant model of retinitis pigmentosa, the injection of RdCVF maintains cone function (as measured by photopic ERG recordings) and the structure of cone outer segments (Yang et al., 2009). Since the rate of cone

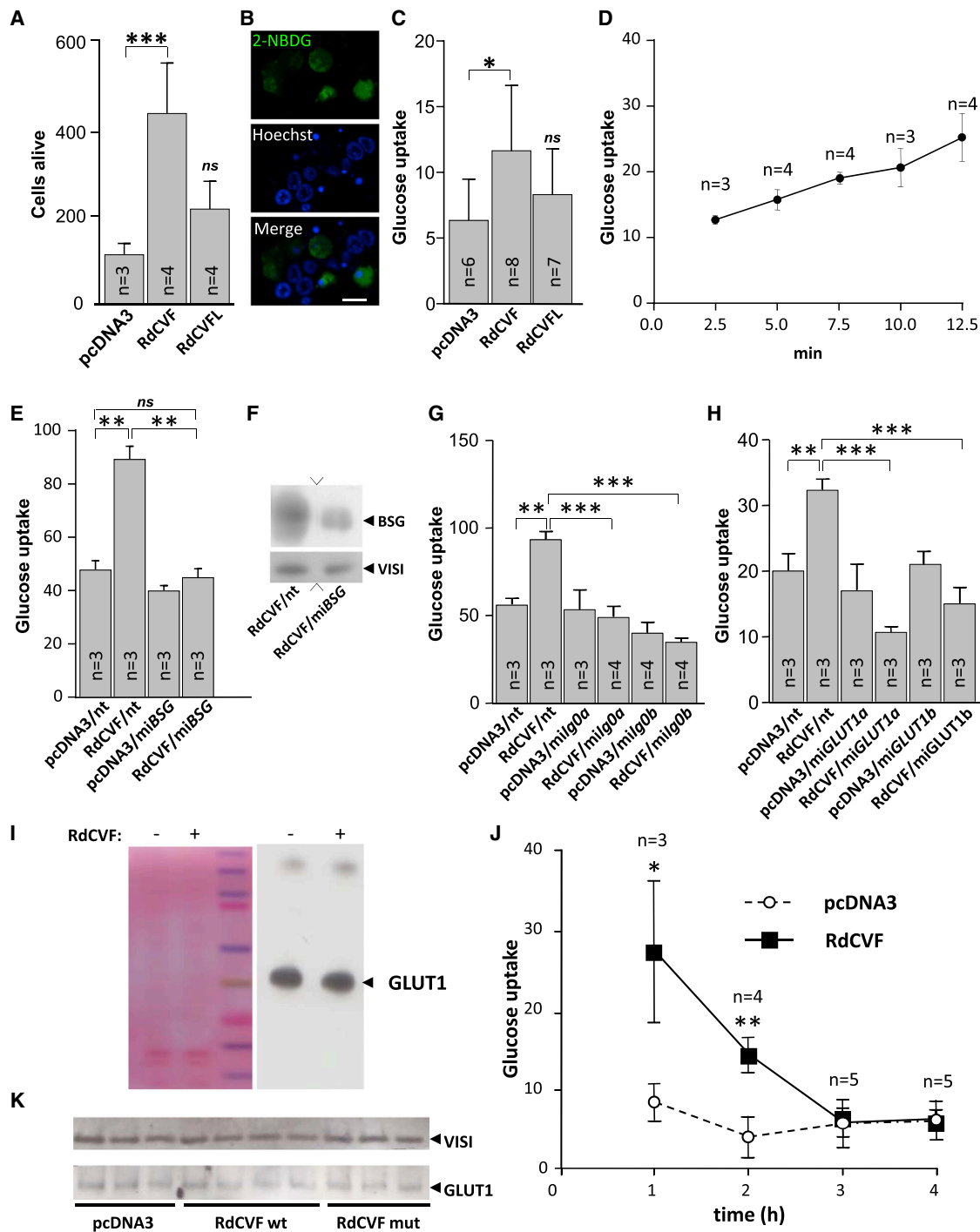


Figure 4. RdCVF Stimulates Glucose Uptake into Cones through Its Interaction with the BSG1-GLUT1 Complex

(A) Effects of conditioned media (CM) from COS-1-transfected cells on survival of cone-enriched cultures. Dunnett's test, SD, $n = 3/4$.

(B) Image of the cells after 2-NBDG uptake, scale bar, 10 μ m.

(C) Effects of conditioned media from COS-1-transfected cells on glucose uptake by cone-enriched culture cells. Dunnett's test, SD, $n = 6-8$.

(D) Kinetic analysis of glucose uptake by cone-enriched culture cells. $n = 3/4$.

(E) Effect of silencing basigin in cone-enriched cultures on RdCVF-mediated cell stimulation of glucose uptake. nt, non-targeting siRNA. Tukey's test, SEM, $n = 3$.

(F) Western blotting analysis of basigin expression in cone-enriched cultures after basigin silencing.

(G) Effect of silencing basigin-1 in cone-enriched cultures on RdCVF-mediated cell stimulation of glucose uptake. milg0a and milg0b, two siRNA targeting two distinct sequences encoding Ig0, a sequence presents in basigin-1 but not in basigin-2. Dunnett's test, SEM, $n = 3/4$.

(legend continued on next page)

degeneration in the *rd1* mouse is rapid, preventing functional analysis, we sectioned retinas from a group of treated animals at PN22 and measured the length of the cone outer segments after labeling with a mixture of MW and S-opsin antibodies (Figures 5K and S4F). The length of these segments of the *rd1* mice treated with AAV-GFP-2A-RdCVF wt was significantly longer than those of animals injected with AAV-GFP-2A-RdCVF mut and to non-injected animals (0) (Figure 5L). Unexpectedly, the length of the cone outer segments of the mice treated with AAV-GFP was higher than to AAV-GFP-2A-RdCVF mut.

RdCVF Stimulates Aerobic Glycolysis

What is the fate of the RdCVF-mediated glucose entering the cones, and what is the mechanism leading to cone survival? To further explore this pathway, we first manipulated the concentration of glucose in the culture medium of cone-enriched cultures, which demonstrated that glucose induces cone survival (Figure 6A). Using this paradigm, we showed that the fold increase in RdCVF-mediated cone survival was higher at 30 than at 15 mM glucose (Figures 6A and 6B), further strengthening the link between RdCVF and glucose uptake. Increasing lactate concentration has no effect on cone-enriched culture survival (Figure S5D). The concentration of intracellular ATP, presumably produced through the glucose metabolism, was found to be higher in RdCVF-treated cone-enriched culture cells (Figure 6C). The limited amplitude in that effect led us to speculate that ATP increase may be produced by aerobic glycolysis. In this alternative usage of glucose, pyruvate produced by glycolysis is not transported to the mitochondria, but rather metabolized to lactate by lactate dehydrogenase (Figure S5A). This phenomenon, known as the Warburg effect, occurs in cancer cells (Vander Heiden et al., 2009). Treating the cone-enriched culture with a lactate dehydrogenase inhibitor, oxamate, abolished RdCVF-mediated survival without generating any toxic effect on unstimulated cells (Figure 6D). Correspondingly, inhibiting the transport of pyruvate to the mitochondria using UK5099, an inhibitor of mitochondrial pyruvate carrier (MPC), did not prevent the action of RdCVF on cone-enriched cultures (Figure 6E). Glucose can be redirected to the pentose phosphate pathway to produce redox power through the production of NADPH, a cofactor of thioredoxin reductases (Anastasiou et al., 2011). 6-Aminonicotinamide (6-AN), a pentose phosphate pathway inhibitor, did not significantly modulate the effect of RdCVF at the maximal non-toxic dose (Figure 6F). RdCVF did not induce a switch from oxidative phosphorylation to aerobic glycolysis since the unstimulated cone-enriched culture cells metabolize glucose through aerobic glycolysis (Figure 6G). The proton production rate (PPR), which is due to the lactate transported in the culture medium through the lactate transporter of the cells, was increased following glucose injection, while the oxygen con-

sumption rate (OCR) resulting from oxidative phosphorylation was slightly reduced. Oligomycin, an inhibitor of mitochondrial ATP synthase, forces the glucose into the aerobic glycolysis pathway. The addition of oxamate reduced the proton production rate prior or after glucose addition. The extracellular acidification rate (ECAR) induced by glucose is inhibited by 2-NBDG, the non-metabolized analog of glucose (Figure 6H).

In order to translate this observation to cones in their natural environment, we examined the expression of a marker of aerobic glycolysis, hexokinase 2 (HK2), in the mouse retina (Wolf et al., 2011). The analysis of the retinal transcriptome of the wt mouse showed an increase in *Hk2* expression that parallels the post-natal maturation of photoreceptors from PN9 to PN21 (Figure 6I). The degeneration of rods in the *rd1* retina prevented a rise in *Hk2* expression most likely because of the death of rods in that model. We also examined the expression of HK1 and HK2, the two major hexokinases involved in glucose metabolism by western blotting. HK2 was expressed at higher levels in the retina than in the brain of wt mice at PN21 and PN35 (Figure 6J). The *rd1* retina is rod-less by PN21, and it is at this age that cones degenerate (Figure S6A). HK2 expression was reduced in the *rd1* retina at PN21 and further reduced by PN35, as if expressed by both rods and cones. The expression of HK1 was not affected by photoreceptor maturation or degeneration. Layers of the wt retina were isolated by vibratome sectioning to directly analyze expression of HK1 and HK2 in the outer retina (OR) containing the photoreceptors and in the inner retina (IR) devoid of photoreceptors (Clérin et al., 2014). HK2 was found almost exclusively in the outer retina layer confirming that this marker of aerobic glycolysis is located in photoreceptors themselves as observed by others (Reidel et al., 2011) (Figure 6K). HK1 was found mostly in the inner retina.

Absence of Action of RdCVF on Rods In Vitro

We examined the expression of basigin (BSG1 + BSG2) in the mature retina (PN27) of the wt mouse by immunohistochemistry. The major site of basigin expression was the retinal pigmented epithelium (Daniele et al., 2008), but a specific signal was co-localized with ATP1A, a marker of the inner segment of photoreceptors (Figure 7A). GLUT1 was widely expressed across the entire retina, including the retinal pigmented epithelium and the inner segment of photoreceptors, as previously observed (Gospe et al., 2010) (Figure 7B). We next examined flat-mounted retinas from PN35 rod-less *rd1* mice (Figures 7C and 7D). In the absence of rods, basigin staining does not colocalize but is associated with that of PNA, a marker of the cone matrix sheath that seems to surround basigin labeling. Western blotting was used to distinguish basigin-1 from basigin-2 through their migration. At PN21, one of the two major products (H) detected in wt mouse retina was reduced in *rd1* retina, while the other product (L) was unchanged (Figure 7E). The expression

(H) Effect of silencing GLUT1 in cone-enriched cultures on RdCVF-mediated cell stimulation of glucose uptake. miGLUT1a and miGLUT1b, two siRNA targeting two distinct sequences encoding GLUT1. Dunnett's test, SEM, $n = 3$.

(I) Expression of GLUT1 at the surface of cone-enriched culture cells in the presence of RdCVF. A non-permeable crosslinking reagent was used to purify cell-surface proteins from cone-enriched culture cells. Left: Ponceau staining.

(J) Kinetics of glucose uptake by cone-enriched culture cells after removing RdCVF stimulation. Student test, SD, $n = 3-5$.

(K) Western blotting analysis of the expression of VIS1 and GLUT1 by cone-enriched cultures in the presence of conditioned media from COS-1-transfected cells. $n = 3/4$.

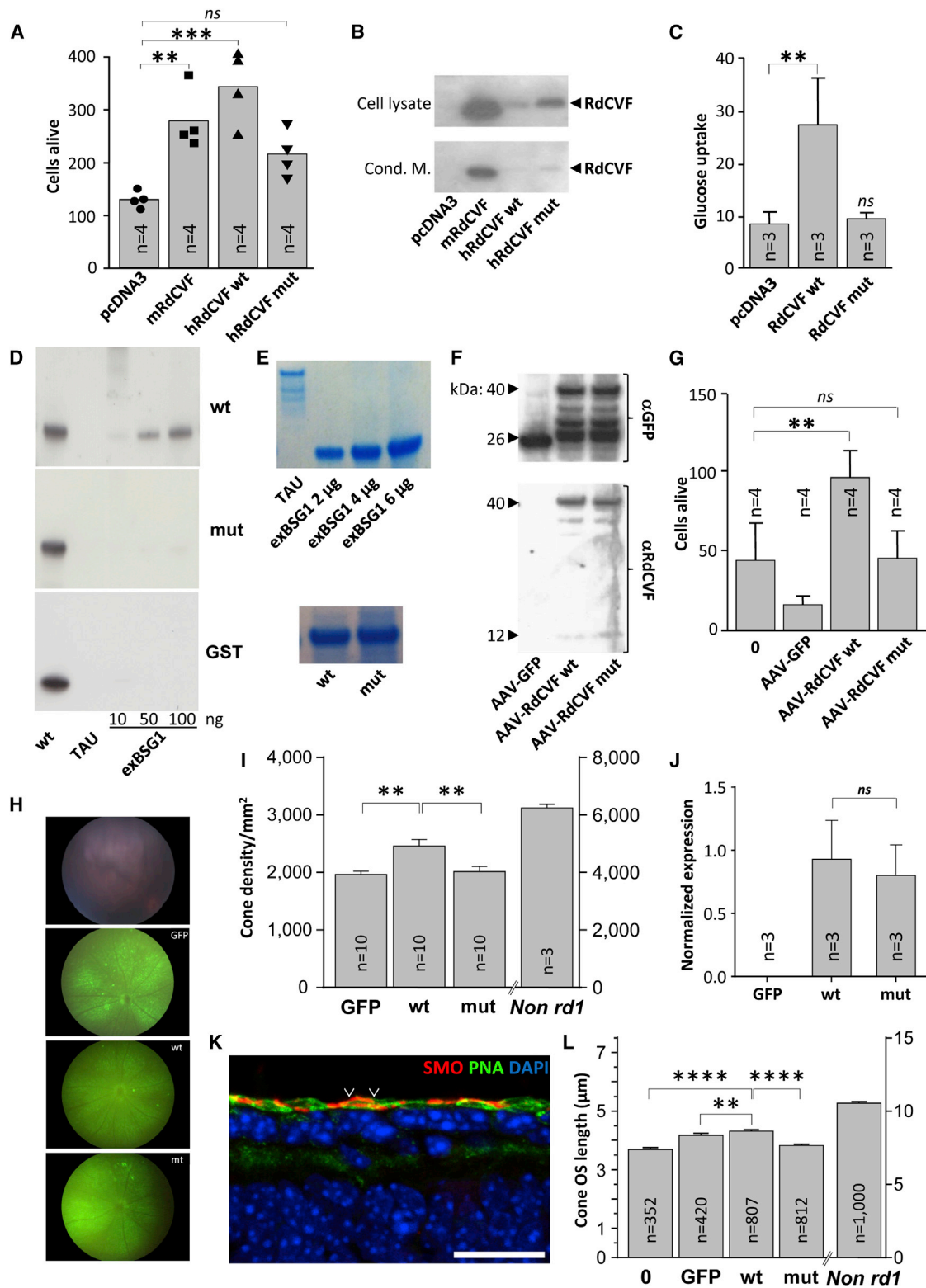


Figure 5. The E64K Mutant of Human RdCVF Is Not Able to Sustain Cone Survival

(A) Effects of conditioned media from COS-1-transfected cells on survival of cone-enriched cultures. mRdCVF and hRdCVF, mouse and human RdCVF, respectively. Wt and mut, 64E and 64K, respectively. Dunnett's test, SD, n = 4.

(B) Western blotting analysis of RdCVF expression in COS-1-transfected cells. Cond. M., conditioned media.

(legend continued on next page)

of GLUT1 remained stable in both genotypes. Following isolation of specific layers by vibratome sectioning, we observed that the H and L products were mainly found in the outer retina and the inner retina layers, respectively (Figure 7F). The absence of rhodopsin in the inner retina showed that there is no major photoreceptors contamination in the inner retina fraction. GLUT1 is expressed at similar level in the outer and the inner retina (Figure 7G).

Basigin-1 is heavily glycosylated (Bai et al., 2014). We deglycosylated mouse retinal extracts to reveal the polypeptidic nature of H and L. Using wt retinal extracts at PN21 and PN27, we showed that H and L correspond, respectively, to basigin-1 and basigin-2 (Figure 7H). In the *rd1* retina, basigin-1, but not basigin-2, was reduced by photoreceptor degeneration.

We then examined the expression of basigin in isolated rods and cones. We isolated photoreceptor outer segments and part of the inner segment by vortexing wt retina (Jaillard et al., 2012) and found that basigin (BSG1 + BSG2) was expressed by both cones and rods (Figure 7I). Basigin localized to outer and inner segments of rods. The expression of the RdCVF receptor by rods suggests that RdCVF might also protect rods by autocrine signaling. We have previously shown that RdCVF has no protective activity toward rods in two models of RP, the *rd1* mouse and the P23H rat; however, these rods express a mutation (Léveillard et al., 2004; Yang et al., 2009). We prepared pure cultures of wt photoreceptors (95% rods) by sectioning of wt mouse retina (*Nxn1*^{+/+}) at PN8 (Clérin et al., 2014). The cultures were incubated with conditioned medium from COS-1 cells transfected with RdCVF, and the number of living cells was counted 5 days later. Fetal calf serum (FCS) increased the number of living cells showing that rods were degenerating over this period (Figure 7J). We did not observe any protective effect of RdCVF on these cultures but reasoned that this could result from the endogenous expression of RdCVF. However, this effect was ruled out since similar results were obtained using *Nxn1*^{-/-} cultures (Figure 7K). The main difference observed was a large reduction in numbers of living cells in absence of *Nxn1*. We showed that cultured rods express basigin (Figures 7L, S6A, and S6B). In extracts prepared from these cells (R), we detected basigin-1 expression by western blotting (Figure 7M). When studying basigin-1 location in cone-enriched cultures and mouse rod cultures, we found that basigin-1 is mainly located in the membrane fraction in both species (Figure 7N). Finally, we studied the effect of glucose on cell viability. Raising the concentra-

tion of glucose from 25 to 50 mM did not increase cell survival in these cultures (Figure 7O). The protective effect of fetal calf serum was only partially blocked by oxamate.

RdCVF Cell-Surface Receptor in Human Retina

To study RdCVF signaling in the human retina, we generated polyclonal antibodies against two distinct peptides of human Ig0. Antibodies BSG1a and BSG1b specifically recognized the H band and the deglycosylated 42 kDa polypeptide in human retinal extract (Figure 7P). A human ocular globe was dissected, and four retinal punches were collected from the fovea and increasingly peripheral eccentricities. The absence of rhodopsin (RHO) and the presence of cone arrestin (ARR3) in specimen 1 showed that it corresponds to the fovea (Figure 7Q). α BSG and α BSG1a detected basigin-1 expression of in all retinal specimens, including the fovea. Basigin-1 expression was also analyzed in sections of normal and RP retinas. In normal human retina, basigin-1 expression was located in the photoreceptor layer, matching the pattern observed with the anti-L/MW antibodies (Figure 7R). In macaque retina, basigin-1 expression localized to inner segments and outer segments of cones, identified by their unique morphology (Figures 7S and S6C). Interestingly, in the advanced stage RP, with no remaining rods, some positive cells that may correspond to cones without outer segment were observed (Figure 7T).

DISCUSSION

RdCVF Accelerates GLUT1 Transport Function

The truncation within the thioredoxin fold in RdCVF removes the region of the protein that interacts with thioredoxin reductases and recycles enzyme activity (Holmgren, 1985). Therefore, the identification of a cell-surface receptor mediating the trophic action of RdCVF on target cells is not surprising. RdCVF is not an enzyme, but its sequence encompasses the conserved dithiol catalytic site CXXC. Oxidative stress induces the secretion of both TRX80 and thioredoxin-1 (Sahaf and Rosen, 2000), and the redox-active site in thioredoxins is essential for the release of extracellular thioredoxin-1 in response to H₂O₂ (Kondo et al., 2004). By analogy, we propose that RdCVF is secreted in an oxidized form that could react with the free cysteine of basigin-1 or GLUT1, resulting in disulfide bridge formation, which may constitute the triggering mechanism activating the pathway (Matsuo and Yodoi, 2013). There is a free surface

(C) Effects of conditioned media from COS-1-transfected cells on glucose uptake by cone-enriched culture cells. Dunnett's test, SD, n = 3.

(D) Far-western blotting analysis of the purified extracellular domain of human basigin-1 (exBSG1) using GST, GST-RdCVF wt (wt), and GST-RdCVF mut (mut).

(E) Coomassie-stained gel of purified TAU, exBSG1, GST-RdCVF WT, and GST-RdCVF mut.

(F) Expression of human RdCVF in conditioned media from COS-1-transfected cells with plasmids pAAV-GFP, pAAV-GFP-2A-RdCVF WT, and pAAV-GFP-2A-RdCVF mut.

(G) Effects of conditioned media from COS-1-transfected cells with pAAV-GFP, pAAV-GFP-2A-RdCVF WT, and pAAV-GFP-2A-RdCVF mut on survival of cone-enriched cultures. Dunnett's test, SD, n = 4.

(H) Ocular fundus imaging from AAV-treated *rd1* mice prior to sacrifice.

(I) Cone density of treated *rd1* mice at PN49. Mann-Whitney test, SEM, n = 3/10.

(J) qRT-PCR analysis of RdCVF mRNA expression in the treated *rd1* mouse retina. Dunnett's test, SD, n + 3.

(K) Immunohistochemical analysis of the cone outer segments of treated *rd1* mice. Red, S- and M-opsin antibodies (SMO); green, PNA; blue, DAPI. White arrowheads delimit a cone outer segment. Scale bar, 16 μ m.

(L) Cone outer segment length of treated *rd1* mice at PN22. 0, non-injected mice. Mann-Whitney test, SEM, n = 352/420/807/1,000.

See also Figure S4.

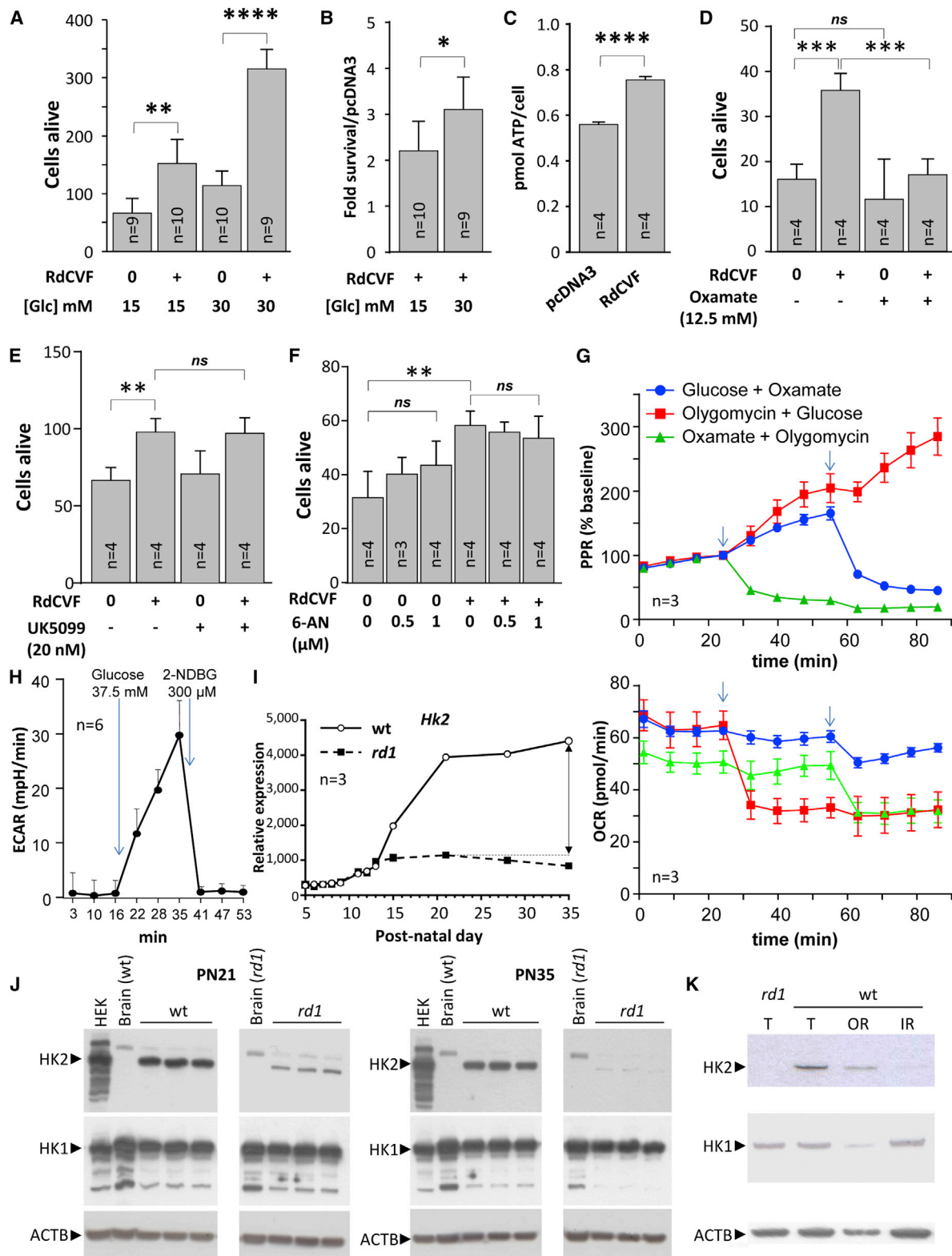


Figure 6. RdCVF Survival Effect Relies on the Stimulation of Aerobic Glycolysis

(A) Effect of glucose (Glc) concentration on survival of cone-enriched culture cells. Student test, SD, n = 9/10.

(B) Effect of glucose concentration on RdCVF-mediated survival of cone-enriched culture cells. Student test, SD, n = 9/10.

(legend continued on next page)

exposed cysteine in human IgG, although it is not conserved in chicken (Redzic et al., 2011).

Alternatively, conformational changes induced by RdCVF may drive activation. GLUT1 catalyzes the rate-limiting step in supplying cells of the CNS. There are, to our knowledge, no extracellular signals known to regulate the activity of GLUT1 when it is located at the surface of the cell. GLUT1 exists in equilibrium between its homodimeric and homotetrameric quaternary structure (De Zutter et al., 2013). Each subunit of GLUT1 contains an extracellular disulfide bridge (C347 and C421) that stabilizes the tetrameric structure and thereby accelerates transport function by increasing the V_{max} of transport and decreasing the K_m (Hebert and Carruthers, 1992). GLUT1 reduction causes GLUT1 tetramers to dissociate into dimers. RdCVF binding to basigin-1 may somehow displace the equilibrium toward the tetramer, accelerating GLUT1 transport function and stimulating glucose uptake by cones. This model is consistent with our data showing that increased glucose uptake simultaneously requires the presence of RdCVF, basigin-1, and GLUT1 (Figure 4J).

RdCVF Stimulates Aerobic Glycolysis

When Otto Warburg described aerobic glycolysis as a hallmark of cancer cells, he also identified the retina as an exception (Warburg, 1956). The current hypothesis for the existence of the aerobic glycolysis in mammalian retina is linked to glucose metabolism involved in the daily renewal of the outer segments of photoreceptors (10% daily). The high content of outer segment polyunsaturated fatty acids results in rapid lipid peroxidation through photooxidation by incident light. Outer segments' renewal involves daily shedding of distal outer segment tips, their phagocytosis by the adjacent retinal pigmented epithelial cells, and their renewal from the photoreceptor inner segment. In adult retina, the photoreceptors maintain a constant outer segment length by balancing the shedding of discs and the assembly of new discs (Young and Bok, 1969). Outer segment renewal is energetically demanding because proteins and lipids, necessary to build outer segments, have to be synthesized at high rate in the inner segments (Casson et al., 2013). Similarly, cancer cells proliferate and rely on the production of carbohydrate intermediates at a high rate (Vander Heiden et al., 2009). Metabolic reprogramming of cancer cells to use the Warburg effect is a primary transformation event (Sebastián et al., 2012). The observation that cone precursors are prone to tumorigenesis may be related to our observation (Xu et al., 2014) (Figure 6G). The splice isoforms of the pyruvate kinase gene associated with aerobic glycolysis is also expressed by photoreceptors (Lindsay et al., 2014).

Interestingly, when RdCVF was injected in the P23H rat model, we observed that the cone outer segments were longer than in controls (Yang et al., 2009). The excess of glucose entering cone cells after RdCVF stimulation may not be converted entirely into lactate. A portion of this glucose is probably converted into lipid precursors for cone outer segment renewal via dihydroxyacetone phosphate (DHAP) (Figure S5A). Our results are in agreement with the protective effect of insulin on cones of the *rd1* mouse since insulin is driving the GLUT4 to the cell surface (Punzo et al., 2009). Lactate should clear out from the interphotoreceptor space through the retinal pigmented epithelium, since MCT, similarly to GLUT1, is a facilitated transporter that carries its substrate along its concentration gradient. Increased lactate concentration in interphotoreceptor space and consequently a decrease of aerobic glycolysis are likely responsible for photoreceptor malfunction in the *Mct3*^{-/-} mouse (Daniele et al., 2008). The finding that *MCT3* locus (*SLC16A8*) is genetically associated with age-related macular degeneration (Fritsche et al., 2013) suggests a possible relationship between *NXNL1* and age-related macular degeneration.

Cones versus Rods

Basigin-1 is expressed at the surface of rods and cones (Figures 7N and S5C). We have not identified significant differences in the expression of glycolytic genes between cones and rods that could explain the cone-specific protective properties of RdCVF (Figure S5B). However, the absence of protection of rods by RdCVF in vitro may be related to the fact that in contrast to cones, rod survival is not stimulated by increased glucose in the culture medium (Figures 6A and 7O). The mechanistic explanation for the rod-cone difference is presently unknown. RdCVF does not protect rods in models of RP, but the mutations are potentially dominant over any trophic activity of RdCVF. Non-cell-autonomous rod degeneration was observed in *rd/s* mosaic mutant male mice carrying a rescue transgene integrated into the X chromosome (Kedzierski et al., 1998). Eventually, RdCVF protective activity on rods could be revealed using this model.

RdCVF Signaling Translated into Medical Practice

In RP patients, cone outer segments are shortened with the progression of the disease (Mitamura et al., 2013), although cones seem to survive even in advanced cases of RP. We have obtained preliminary evidence for the expression of basigin-1 in surviving cones in the RP retina (Figure 7T). The mechanism of action revealed here implies that administration of RdCVF in patients suffering from RP could not only stabilize central vision

(C) Concentration of ATP in cone-enriched culture cell extract in the presence of conditioned media from COS-1-transfected cells. Student test, SD, $n = 4$.

(D) Effects of the lactate dehydrogenase inhibitor, oxamate, on RdCVF-mediated survival of cone-enriched cultures. Dunnett's test, SD, $n = 4$.

(E) Effects of the mitochondrial pyruvate carrier inhibitor, UK5099, on RdCVF-mediated survival of cone-enriched cultures. Student test, SD, $n = 4$.

(F) Effects of the inhibitor of the pentose phosphate pathway, 6-AN, on RdCVF-mediated survival of cone-enriched cultures. Student test, SD, $n = 3/4$.

(G) Proton production rate (PPR) and oxygen consumption rate (OCR) of cone-enriched culture cells. Arrows indicate injection time. $n = 3$.

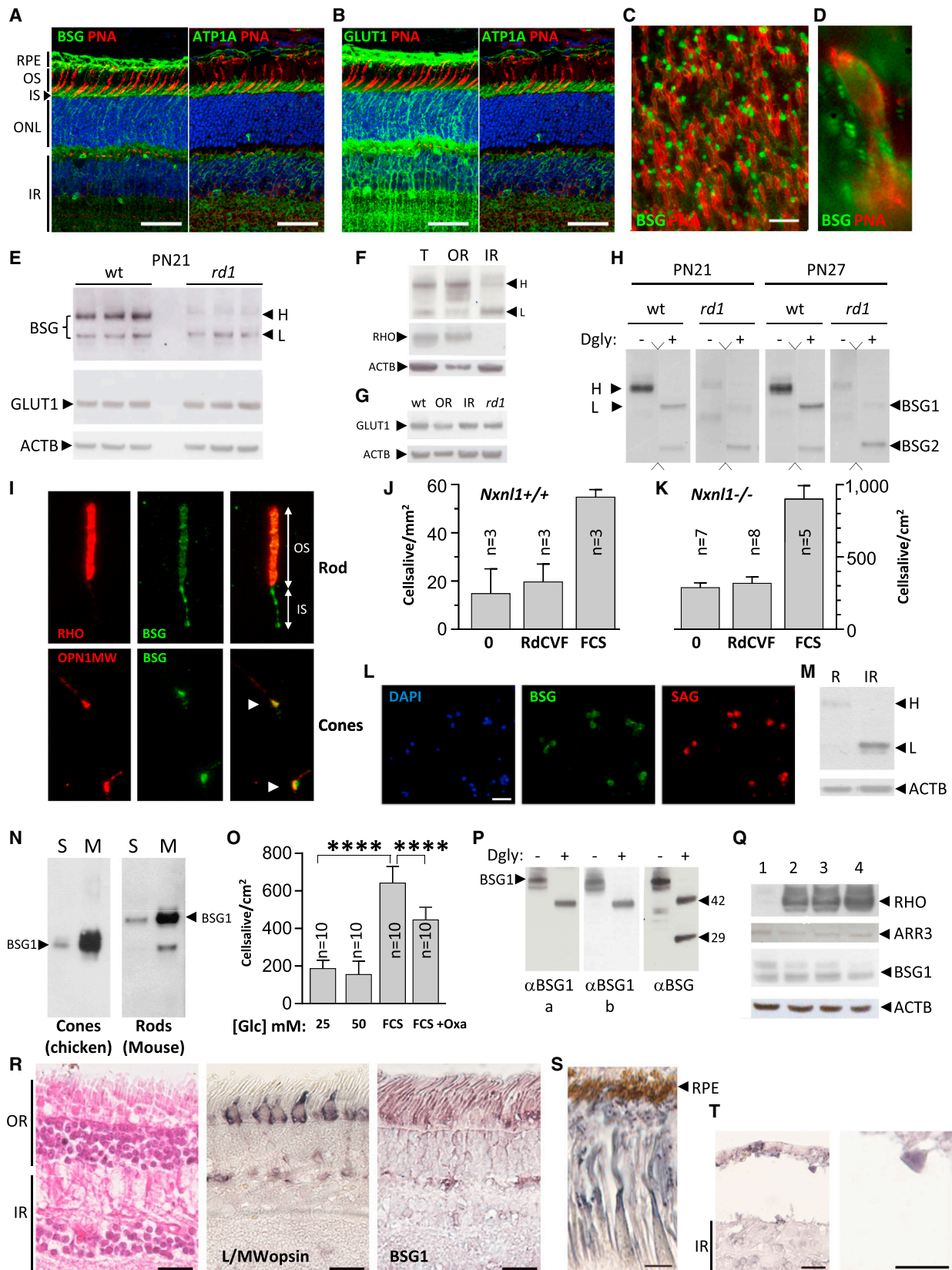
(H) Effect of 2-NBDG of extracellular acidification rate (ECAR). Arrows indicate injection time. $n = 6$.

(I) Expression of the mRNA of hexokinase 2 (*Hk2*) during the maturation of photoreceptors in the wild-type (wt) and *rd1* mouse retina. $n = 3$.

(J) Western blotting analysis of HK1 and HK2 expression in the WT and *rd1* retina at post-natal day 21 (PN21) and PN35. ACTB, cytoplasmic actin. $n = 3$.

(K) Expression of HK1 and HK2 in the wild-type retinal layers of a WT mouse at PN35 after vibratome sectioning. T, OR, and IR, whole retina, outer retinal layer, and inner retinal layer, respectively.

See also Figure S5A.



(legend on next page)

but also ameliorate cone vision by stimulating cone outer segments re-growth.

EXPERIMENTAL PROCEDURES

All experiments were approved by the UPMC ethical committee (Darwin # Ce5/2011/013) and through the French regulation to conduct animal research (A-75-1863, TL). All experiments were performed in accordance with the ARVO statement for the use of animals in ophthalmic and vision research. Human specimens were obtained under the approval DC-2008-346.

Detailed protocols for each figure panel are provided in [Extended Experimental Procedures](#). Binding, immunohistochemistry, immunocytochemistry, co-immunoprecipitation, western blotting, alkaline phosphatase fusion proteins, RT-PCR, electroporation, transient transfections, RNA silencing, protein extraction and fractionation, recombinant protein production and purification, FRET, glucose uptake, MS/MS analysis, recombinant AAV construction and production, ATP measurement, and respirometry were performed using conventional methods. Cone-enriched culture and pure cultures of mouse rods are described in [Léveillard et al. \(2004\)](#) and [Clérin et al. \(2014\)](#). Outer and inner segments are described in [Jaillard et al. \(2012\)](#). Cone counting in the *rd1* retina ([Clérin et al., 2011](#)) was done as a double-blind experiment, as was measurement of cone outer segment length. Far-western blotting was performed as described ([Léveillard et al., 1996](#)).

Statistical Analysis

Data are expressed as average \pm SD or SEM. Statistical analyses were performed using GraphPad. Data were analyzed by unpaired Student's *t* tests, and, for more than two groups, we used one-way or two-way ANOVA analyses of variance followed by Tukey's or Dunnett's multiple comparison post-tests. Statistical significance is defined as $p < 0.05$ or less and indicated by asterisks.

ACCESSION NUMBERS

The accession number for the transcriptomic data reported in this paper is GEO: GSE62020. The accession number for the proteomic data reported in this paper have been deposited via the PRIDE partner repository to the ProteomeXchange Consortium: PXD001384, DOI <http://dx.doi.org/10.6019/PXD001384>, and PXD001715, DOI <http://dx.doi.org/10.6019/PXD001715>.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, one table and six figures and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.03.023>.

AUTHOR CONTRIBUTIONS

N.A.-A. created [Figures 1D–1G, 2A–2D, 2G, 4C, 4D, 5F, 5H–J, 5L, S1D–S1F, S2D, and S2E](#). R.F. created [Figures 3A–3D, 4E–4H, 5D, 5E, 7O, 7P, and S3A–S3C](#). G.M.-P. created [Figures 1A, 1B, 2B, 2C, 3E, 4A–4D, 4J, 4K, 5A–5C, 6A–6H, S2A, and S2B](#). E.C. created [Figures 5I–5L, 6J, 6K, 7A–7T, S1A, S1B, S2C, S4D–S4F, S6B, and S6C](#). E.M. created [Figures 1B, S1A, and S1B](#). F.D. and A.v.D. performed MS/MS analyses. C.J. performed AAV injections. L.C.B. and D.D. performed AAV-based method production injection. L.P. performed 1C. F. Blond performed bioinformatic analyses. J.B. performed recombinant protein productions and purifications. S.R. performed northern blotting. X.N. performed FRET. F. Bouillaud performed respirometry.

ACKNOWLEDGMENTS

We thank C. Petit, J. Kaplan, J. Ochrietor, D.J. Zack, M. Canto-Soler, A. Chédotal, D. Hicks, R. Goument, N. Berdugo, V. Ferracane, Y. Zagar, M.L. Niepon, G. Lucchi, V. and M. Fradot, M. Desrosiers, G. Tarlet, S. Fouquet, NDRI, cornea bank of Beverwijk, and MIRCEN for scientific advice and/or biological materials. This work was supported by Sanofi, UPMC, Inserm, CNRS, EC, FFB, and ANR. J.-A.S., and T.L. are patent holders on the use of RdCVF to treat inherited retinal degenerations. A.v.D., J.-A.S., and T.L. are patent holders on the use of BSG1 RdCVF to treat inherited retinal degenerations.

Received: October 28, 2014

Revised: January 21, 2015

Accepted: March 5, 2015

Published: May 7, 2015

REFERENCES

Anastasiou, D., Poulogiannis, G., Asara, J.M., Boxer, M.B., Jiang, J.K., Shen, M., Bellinger, G., Sasaki, A.T., Locasale, J.W., Auld, D.S., et al. (2011).

Figure 7. Expression Analysis of BSG1 and the Absence of Trophic Effect of RdCVF on Rod Photoreceptors

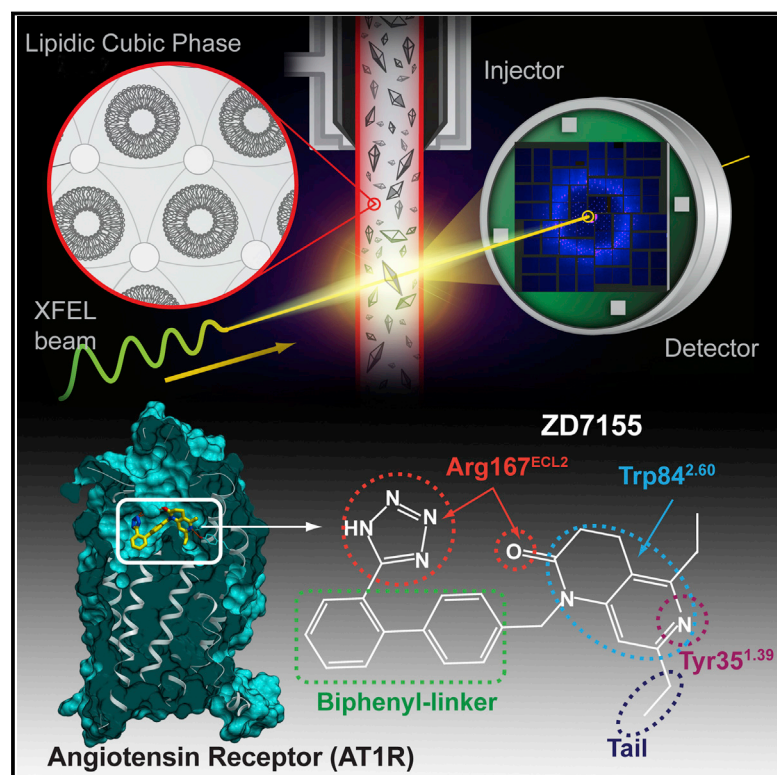
- (A) Immunohistochemical analysis of basigin expression in the retina of a wt mouse at PN26. BSG (BSG1 + BSG2); PNA, peanut agglutinin; ONL, outer nuclear layer.
- (B) Immunohistochemical analysis of GLUT1 expression in the retina of a wt mouse at PN26. Scale bar, 28 μ m.
- (C) Immunohistochemical analysis of basigin expression in the flat-mounted *rd1* retina at PN35.
- (D) Higher magnification.
- (E) Western blotting analysis of basigin (BSG) expression in the wt and *rd1* retina at PN21. ACTB, cytoplasmic actin. $n = 3$.
- (F) Expression of basigin in the retinal layers of a WT mouse at PN35 after vibratome sectioning. T, OR, and IR, whole retina, outer retinal layer, and inner retinal layer, respectively. RHO, rhodopsin.
- (G) Expression of GLUT1 in the retinal layers of a wt mouse at PN35 after vibratome sectioning. wt and *rd1*, PN35 retinas without sectioning.
- (H) Expression of basigin-1 and basigin-2 in the wt and *rd1* retina at PN21 and PN27 after deglycosylation (Dgly.). H and L, high- and low-molecular-weight protein, respectively.
- (I) Immunohistochemical analysis of basigin expression in rod and cone inner (IS) and outer (OS) segments. OPN1MW, media wave cone opsin.
- (J) Effects of conditioned media from COS-1-transfected cells on survival of pure rod cultures from *Nxn1^{+/+}* retina. FCS, fetal calf serum. SD, $n = 3$.
- (K) Results obtained as in (I) from *Nxn1^{-/-}* mice. SD. Notice the difference in the y axis scale between (I) and (J). $n = 5/8$.
- (L) Immunocytochemical analysis of the rod cultures with anti-BSG antibodies. SAG, rod arrestin. Scale bar, 14 μ m.
- (M) Expression of basigin by rod cultures (R) and in inner retina (IR).
- (N) Basigin sub-cellular localization in cone-enriched and mouse neural retina.
- (O) Effect of glucose (Glc) concentration on survival of rod cultured cells. Unpaired *t* test, SD, $n = 10$.
- (P) Rabbit polyclonal antibodies, α BSG1a and α BSG1b, raised against the IgD domain, specific to Basigin-1. Dgly, deglycosylation.
- (Q) Western blotting analysis of the expression of basigin-1 in the human retina from the center (1) to the periphery (4). ARR3, cone arrestin.
- (R) Immunohistochemical analysis of the expression of basigin-1 in normal human retina. Scale bar, 25 μ m. Left, H&E staining.
- (S) Immunohistochemical analysis of the expression of basigin-1 in the macaque retina. Arrowhead, retinal pigmented epithelium. Scale bar, 10 μ m.
- (T) Immunohistochemical analysis of the expression of basigin-1 in the human retinitis pigmentosa retina. Scale bar, 30 μ m.
- See also [Figure S6](#).

- Inhibition of pyruvate kinase M2 by reactive oxygen species contributes to cellular antioxidant responses. *Science* 334, 1278–1283.
- Bai, Y., Huang, W., Ma, L.T., Jiang, J.L., and Chen, Z.N. (2014). Importance of N-glycosylation on CD147 for its biological functions. *Int. J. Mol. Sci.* 15, 6356–6377.
- Byrne, L.C., Dalkara, D., Luna, G., Fisher, S.K., Clérin, E., Sahel, J.A., Léveillard, T., and Flannery, J.G. (2015). Viral-mediated RdCVF and RdCVFL expression protects cone and rod photoreceptors in retinal degeneration. *J. Clin. Invest.* 125, 105–116.
- Casson, R.J., Chidlow, G., Han, G., and Wood, J.P. (2013). An explanation for the Warburg effect in the adult mammalian retina. *Clin. Experiment. Ophthalmol.* 41, 517.
- Chalmel, F., Léveillard, T., Jaillard, C., Lardenois, A., Berdugo, N., Morel, E., Koehl, P., Lambrou, G., Holmgren, A., Sahel, J.A., and Poch, O. (2007). Rod-derived Cone Viability Factor-2 is a novel bifunctional-thioredoxin-like protein with therapeutic potential. *BMC Mol. Biol.* 8, 74.
- Chen, Y., Zhang, J., and Zhang, X.Y. (2015). 2-NBDG as a marker for detecting glucose uptake in reactive astrocytes exposed to oxygen-glucose deprivation in vitro. *J. Mol. Neurosci.* 55, 126–130.
- Clérin, E., Wicker, N., Mohand-Saïd, S., Poch, O., Sahel, J.A., and Léveillard, T. (2011). e-conome: an automated tissue counting platform of cone photoreceptors for rodent models of retinitis pigmentosa. *BMC Ophthalmol.* 11, 38.
- Clérin, E., Yang, Y., Forster, V., Fontaine, V., Sahel, J.A., and Léveillard, T. (2014). Vibratome sectioning mouse retina to prepare photoreceptor cultures. *J. Vis. Exp.* <http://dx.doi.org/10.3791/51954>.
- Cronin, T., Raffelsberger, W., Lee-Rivera, I., Jaillard, C., Niepon, M.L., Kinzel, B., Clérin, E., Petrosian, A., Picaud, S., Poch, O., et al. (2010). The disruption of the rod-derived cone viability gene leads to photoreceptor dysfunction and susceptibility to oxidative stress. *Cell Death Differ.* 17, 1199–1210.
- Daniele, L.L., Sauer, B., Gallagher, S.M., Pugh, E.N., Jr., and Philp, N.J. (2008). Altered visual function in monocarboxylate transporter 3 (Slc16a8) knockout mice. *Am. J. Physiol. Cell Physiol.* 295, C451–C457.
- De Zutter, J.K., Levine, K.B., Deng, D., and Carruthers, A. (2013). Sequence determinants of GLUT1 oligomerization: analysis by homology-scanning mutagenesis. *J. Biol. Chem.* 288, 20734–20744.
- Delyfer, M.N., Raffelsberger, W., Mercier, D., Korobelnik, J.F., Gaudric, A., Charteris, D.G., Tadayoni, R., Metge, F., Caputo, G., Barale, P.O., et al. (2011). Transcriptomic analysis of human retinal detachment reveals both inflammatory response and photoreceptor death. *PLoS ONE* 6, e28791.
- Dryja, T.P., Hahn, L.B., Reboul, T., and Arnaud, B. (1996). Missense mutation in the gene encoding the alpha subunit of rod transducin in the Nougaret form of congenital stationary night blindness. *Nat. Genet.* 13, 358–360.
- Elachouri, G., Lee-Rivera, I., Clérin, E., Argentini, M., Fridlich, R., Blond, F., Ferracane, V., Yang, Y., Raffelsberger, W., Wan, J., et al. (2015). The thioredoxin RdCVFL protects against photo-oxidative retinal damage. *Free Radic. Biol. Med.* 81, 22–29.
- Fridlich, R., Delalande, F., Jaillard, C., Lu, J., Poidevin, L., Cronin, T., Perrocheau, L., Millet-Puel, G., Niepon, M.L., Poch, O., et al. (2009). The thioredoxin-like protein rod-derived cone viability factor (RdCVFL) interacts with TAU and inhibits its phosphorylation in the retina. *Mol. Cell. Proteomics* 8, 1206–1218.
- Fritsche, L.G., Chen, W., Schu, M., Yaspan, B.L., Yu, Y., Thorleifsson, G., Zack, D.J., Arakawa, S., Cipriani, V., Ripke, S., et al. (2013). Seven new loci associated with age-related macular degeneration. *Nat. Genet.* 45, 433–439.
- Gospe, S.M., 3rd, Baker, S.A., and Arshavsky, V.Y. (2010). Facilitative glucose transporter Glut1 is actively excluded from rod outer segments. *J. Cell Sci.* 123, 3639–3644.
- Hanein, S., Perrault, I., Gerber, S., Dollfus, H., Dufier, J.L., Feingold, J., Munich, A., Bhattacharya, S., Kaplan, J., Sahel, J.A., et al. (2006). Disease-associated variants of the rod-derived cone viability factor (RdCVF) in Leber congenital amaurosis. Rod-derived cone viability variants in LCA. *Adv. Exp. Med. Biol.* 572, 9–14.
- Hebert, D.N., and Carruthers, A. (1992). Glucose transporter oligomeric structure determines transporter function. Reversible redox-dependent interconversions of tetrameric and dimeric GLUT1. *J. Biol. Chem.* 267, 23829–23838.
- Holmgren, A. (1985). Thioredoxin. *Annu. Rev. Biochem.* 54, 237–271.
- Hori, K., Katayama, N., Kachi, S., Kondo, M., Kadomatsu, K., Usukura, J., Muramatsu, T., Mori, S., and Miyake, Y. (2000). Retinal dysfunction in basigin deficiency. *Invest. Ophthalmol. Vis. Sci.* 41, 3128–3133.
- Jaillard, C., Mouret, A., Niepon, M.L., Clérin, E., Yang, Y., Lee-Rivera, I., Aït-Ali, N., Millet-Puel, G., Cronin, T., Sedmak, T., et al. (2012). Nxn12 splicing results in dual functions in neuronal cell survival and maintenance of cell integrity. *Hum. Mol. Genet.* 21, 2298–2311.
- Kedzierski, W., Bok, D., and Travis, G.H. (1998). Non-cell-autonomous photoreceptor degeneration in rds mutant mice mosaic for expression of a rescue transgene. *J. Neurosci.* 18, 4076–4082.
- Kirk, P., Wilson, M.C., Hedde, C., Brown, M.H., Barclay, A.N., and Halestrap, A.P. (2000). CD147 is tightly associated with lactate transporters MCT1 and MCT4 and facilitates their cell surface expression. *EMBO J.* 19, 3896–3904.
- Kondo, N., Ishii, Y., Kwon, Y.W., Tanito, M., Horita, H., Nishinaka, Y., Nakamura, H., and Yodoi, J. (2004). Redox-sensing release of human thioredoxin from T lymphocytes with negative feedback loops. *J. Immunol.* 172, 442–448.
- Léveillard, T., Andera, L., Bissonnette, N., Schaeffer, L., Bracco, L., Egly, J.M., and Wasyluk, B. (1996). Functional interactions between p53 and the TFIIF complex are affected by tumour-associated mutations. *EMBO J.* 15, 1615–1624.
- Léveillard, T., Mohand-Saïd, S., Lorentz, O., Hicks, D., Fintz, A.C., Clérin, E., Simonutti, M., Forster, V., Cavusoglu, N., Chalmel, F., et al. (2004). Identification and characterization of rod-derived cone viability factor. *Nat. Genet.* 36, 755–759.
- Lindsay, K.J., Du, J., Sloat, S.R., Contreras, L., Linton, J.D., Turner, S.J., Sadilek, M., Satrustegui, J., and Hurley, J.B. (2014). Pyruvate kinase and aspartate-glutamate carrier distributions reveal key metabolic links between neurons and glia in retina. *Proc. Natl. Acad. Sci. USA* 111, 15579–15584.
- Matsuo, Y., and Yodoi, J. (2013). Extracellular thioredoxin: a therapeutic tool to combat inflammation. *Cytokine Growth Factor Rev.* 24, 345–353.
- Mitamura, Y., Mitamura-Aizawa, S., Katome, T., Naito, T., Hagiwara, A., Kumagai, K., and Yamamoto, S. (2013). Photoreceptor impairment and restoration on optical coherence tomographic image. *J. Ophthalmol.* 2013, 518170.
- Ochriator, J.D., Moroz, T.P., van Ekeris, L., Clamp, M.F., Jefferson, S.C., deCarvalho, A.C., Fadool, J.M., Wistow, G., Muramatsu, T., and Linser, P.J. (2003). Retina-specific expression of 5A11/Basigin-2, a member of the immunoglobulin gene superfamily. *Invest. Ophthalmol. Vis. Sci.* 44, 4086–4096.
- Pekkari, K., Avila-Cariño, J., Gurunath, R., Bengtsson, A., Scheynius, A., and Holmgren, A. (2003). Truncated thioredoxin (Trx80) exerts unique mitogenic cytokine effects via a mechanism independent of thiol oxidoreductase activity. *FEBS Lett.* 539, 143–148.
- Punzo, C., Kornacker, K., and Cepko, C.L. (2009). Stimulation of the insulin/mTOR pathway delays cone death in a mouse model of retinitis pigmentosa. *Nat. Neurosci.* 12, 44–52.
- Redzic, J.S., Armstrong, G.S., Isern, N.G., Jones, D.N., Kieft, J.S., and Eisenmesser, E.Z. (2011). The retinal specific CD147 IgO domain: from molecular structure to biological activity. *J. Mol. Biol.* 411, 68–82.
- Reichman, S., Kalathur, R.K., Lambard, S., Aït-Ali, N., Yang, Y., Lardenois, A., Ripp, R., Poch, O., Zack, D.J., Sahel, J.A., and Léveillard, T. (2010). The homeobox gene CHX10/VSX2 regulates RdCVF promoter activity in the inner retina. *Hum. Mol. Genet.* 19, 250–261.
- Reidel, B., Thompson, J.W., Farsiu, S., Moseley, M.A., Skiba, N.P., and Arshavsky, V.Y. (2011). Proteomic profiling of a layered tissue reveals unique glycolytic specializations of photoreceptor cells. *Mol. Cell. Proteomics* 10, M110.002469.
- Rubartelli, A., Bajetto, A., Allavena, G., Wollman, E., and Sitia, R. (1992). Secretion of thioredoxin by normal and neoplastic cells through a leaderless secretory pathway. *J. Biol. Chem.* 267, 24161–24164.

- Sahaf, B., and Rosen, A. (2000). Secretion of 10-kDa and 12-kDa thioredoxin species from blood monocytes and transformed leukocytes. *Antiox. Redox* 2, 717–726.
- Schwertassek, U., Balmer, Y., Gutscher, M., Weingarten, L., Preuss, M., Engelhard, J., Winkler, M., and Dick, T.P. (2007). Selective redox regulation of cytokine receptor signaling by extracellular thioredoxin-1. *EMBO J.* 26, 3086–3097.
- Sebastián, C., Zwaans, B.M., Silberman, D.M., Gymrek, M., Goren, A., Zhong, L., Ram, O., Truelove, J., Guimaraes, A.R., Toiber, D., et al. (2012). The histone deacetylase SIRT6 is a tumor suppressor that controls cancer metabolism. *Cell* 151, 1185–1199.
- Vander Heiden, M.G., Cantley, L.C., and Thompson, C.B. (2009). Understanding the Warburg effect: the metabolic requirements of cell proliferation. *Science* 324, 1029–1033.
- Warburg, O. (1956). On the origin of cancer cells. *Science* 123, 309–314.
- Wolf, A., Agnihotri, S., Micallef, J., Mukherjee, J., Sabha, N., Cairns, R., Hawkins, C., and Guha, A. (2011). Hexokinase 2 is a key mediator of aerobic glycolysis and promotes tumor growth in human glioblastoma multiforme. *J. Exp. Med.* 208, 313–326.
- Wright, A.F. (1997). A searchlight through the fog. *Nat. Genet.* 17, 132–134.
- Xu, S.Z., Sukumar, P., Zeng, F., Li, J., Jairaman, A., English, A., Naylor, J., Ciurtea, C., Majeed, Y., Milligan, C.J., et al. (2008). TRPC channel activation by extracellular thioredoxin. *Nature* 451, 69–72.
- Xu, X.L., Singh, H.P., Wang, L., Qi, D.-L., Poulos, B.K., Abramson, D.H., Jhanwar, S.C., and Cobrinik, D. (2014). Rb suppresses human cone-precursor-derived retinoblastoma tumours. *Nature* 514, 385–388.
- Yang, Y., Mohand-Said, S., Danan, A., Simonutti, M., Fontaine, V., Clerin, E., Picaud, S., Leveillard, T., and Sahel, J.A. (2009). Functional cone rescue by RdCVF protein in a dominant model of retinitis pigmentosa. *Mol. Ther.* 17, 787–795.
- Young, R.W., and Bok, D. (1969). Participation of the retinal pigment epithelium in the rod outer segment renewal process. *J. Cell Biol.* 42, 392–403.

Structure of the Angiotensin Receptor Revealed by Serial Femtosecond Crystallography

Graphical Abstract



Authors

Haitao Zhang, Hamiyet Unal, ..., Sadashiva S. Karnik, Vadim Cherezov

Correspondence

cherezov@usc.edu

In Brief

Structure determination of human Angiotensin II type 1 receptor bound to an antagonist using serial femtosecond crystallography with X-ray free-electron laser and docking studies of other common anti-hypertensive drugs into the structure offer insights into design of blood pressure modulators.

Highlights

- Crystal structure of the human Angiotensin II type 1 receptor at 2.9-Å resolution
- Structure is solved by X-ray laser serial femtosecond crystallography
- Antagonist ZD7155 forms critical interactions with Tyr35, Trp84 and Arg167
- Docking reveals binding modes of common angiotensin receptor blockers



Structure of the Angiotensin Receptor Revealed by Serial Femtosecond Crystallography

Haitao Zhang,¹ Hamiyet Unal,² Cornelius Gati,³ Gye Won Han,⁴ Wei Liu,⁵ Nadia A. Zatsepin,⁶ Daniel James,⁶ Dingjie Wang,⁶ Garrett Nelson,⁶ Uwe Weierstall,⁶ Michael R. Sawaya,⁷ Qingping Xu,⁸ Marc Messerschmidt,⁹ Garth J. Williams,¹⁰ Sébastien Boutet,¹⁰ Oleksandr M. Yefanov,³ Thomas A. White,³ Chong Wang,¹¹ Andrii Ishchenko,⁴ Kalyan C. Tirupula,² Russell Desnoyer,² Jesse Coe,⁵ Chelsie E. Conrad,⁵ Petra Fromme,⁵ Raymond C. Stevens,^{1,4,12} Vsevolod Katritch,¹ Sadashiva S. Karnik,² and Vadim Cherezov^{4,*}

¹Department of Biological Sciences, Bridge Institute, University of Southern California, Los Angeles, CA 90089, USA

²Department of Molecular Cardiology, Lerner Research Institute, Cleveland Clinic, Cleveland, OH 44195, USA

³Center for Free Electron Laser Science, Deutsches Elektronen-Synchrotron DESY, 22607 Hamburg, Germany

⁴Department of Chemistry, Bridge Institute, University of Southern California, Los Angeles, CA 90089 USA

⁵Department of Chemistry and Biochemistry, Center for Applied Structural Discovery at the Biodesign Institute, Arizona State University, Tempe, AZ 85287, USA

⁶Department of Physics, Arizona State University, Tempe, AZ 85287, USA

⁷Department of Chemistry and Biochemistry, the UCLA-DOE Institute for Genomics and Proteomics, University of California, Los Angeles, CA 90095, USA

⁸Joint Center for Structural Genomics, Stanford Synchrotron Radiation Light Source, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA

⁹BioXFEL Science and Technology Center, Buffalo, NY 14203, USA

¹⁰Linac Coherent Light Source, SLAC National Accelerator Laboratory, Menlo Park, CA 94025, USA

¹¹Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138, USA

¹²Human Institute, ShanghaiTech University, Shanghai, 201210 China

*Correspondence: cherezov@usc.edu

<http://dx.doi.org/10.1016/j.cell.2015.04.011>

SUMMARY

Angiotensin II type 1 receptor (AT₁R) is a G protein-coupled receptor that serves as a primary regulator for blood pressure maintenance. Although several anti-hypertensive drugs have been developed as AT₁R blockers (ARBs), the structural basis for AT₁R ligand-binding and regulation has remained elusive, mostly due to the difficulties of growing high-quality crystals for structure determination using synchrotron radiation. By applying the recently developed method of serial femtosecond crystallography at an X-ray free-electron laser, we successfully determined the room-temperature crystal structure of the human AT₁R in complex with its selective antagonist ZD7155 at 2.9-Å resolution. The AT₁R-ZD7155 complex structure revealed key structural features of AT₁R and critical interactions for ZD7155 binding. Docking simulations of the clinically used ARBs into the AT₁R structure further elucidated both the common and distinct binding modes for these anti-hypertensive drugs. Our results thereby provide fundamental insights into AT₁R structure-function relationship and structure-based drug design.

INTRODUCTION

Cardiovascular disease remains one of the main causes of death throughout the world despite impressive advances in diagnosis

and therapeutics during the past few decades. Hypertension is the most common modifiable risk factor in cardiovascular disease, as myocardial infarction, stroke, heart failure, and renal disease can be greatly reduced by lowering blood pressure (Zaman et al., 2002). The best known regulator of blood pressure is the renin-angiotensin system (RAS). Over-stimulation of the RAS is implicated in hypertension, cardiac hypertrophy, heart failure, ischemic heart disease, and nephropathy (Balakumar and Jagadeesh, 2014). A cascade of proteolytic reactions in the RAS can generate various angiotensin peptides. Renin cleaves the precursor protein, angiotensinogen, releasing the inactive angiotensin I. Subsequently, angiotensin I is cleaved by angiotensin converting enzyme (ACE) to generate angiotensin II (AngII), angiotensin III, and angiotensin 1-7. These peptides exert diverse functions; angiotensins II and III act as vasoconstrictors, while angiotensin 1-7 acts as a vasodilator (Zaman et al., 2002). AngII is also responsible for cell migration, protein synthesis, endothelial dysfunction, inflammation, and fibrosis (Ramchandran et al., 2006).

In humans, AngII binds to two subtypes of angiotensin G protein-coupled receptors (GPCRs), angiotensin II type 1 receptor (AT₁R) and angiotensin II type 2 receptor (AT₂R) (Oliveira et al., 2007). Almost all physiological and pathophysiological effects of AngII are mediated by AT₁R (de Gasparo et al., 2000), while the function of AT₂R remains largely unknown (Akazawa et al., 2013). AT₁R exhibits multiple active conformations, thereby activating different signaling pathways with differential functional outcomes (Shenoy and Lefkowitz, 2005). The G protein-dependent signaling by AT₁R is vital for normal cardiovascular homeostasis yet detrimental in chronic dysfunction, which associates with cell death and tissue fibrosis and leads to cardiac

hypertrophy and heart failure (Ma et al., 2010). Accumulating evidence suggests that G protein independent β -arrestin mediated signaling by AT₁R confers cardio-protective benefits (Whalen et al., 2011; Wisler et al., 2014).

Targeting the RAS cascade has proven to be effective in the treatment of hypertension, as well as specific cardiovascular and renal disorders. The most commonly used drugs include renin inhibitors, ACE inhibitors, and AT₁R blockers (ARBs). ARBs, or sartans, are non-peptide antagonists and include the well-known anti-hypertensive drugs losartan, candesartan, valsartan, irbesartan, telmisartan, eprosartan, olmesartan, and azilsartan, most of which share a common biphenyl-tetrazole scaffold (Burnier and Brunner, 2000; Imaizumi et al., 2013; Miura et al., 2013a; Miura et al., 2013b). These ARBs are now extensively used for the treatment of cardiovascular diseases, including hypertension, cardiac hypertrophy, arrhythmia, and heart failure. There is additional interest in ARBs regarding their efficacy in the treatment of blood-vessel diseases such as Marfan-like syndrome, aortic dissection, and aortic aneurysms (Keane and Pyritz, 2008; Ramanath et al., 2009).

Previous functional studies on AT₁R have provided numerous clues into AT₁R activation and inhibition mechanisms (Oliveira et al., 2007). Despite its high medical relevance and decades of research, the structure of AT₁R and the binding mode of ARBs, however, are still unknown, which limits our understanding of the structural basis for AT₁R function and modulation and precludes the rational optimization of AT₁R lead compounds. One such experimental anti-hypertensive compound is ZD7155, a high-affinity antagonist and precursor to the anti-hypertensive drug candesartan. ZD7155 has a biphenyl-tetrazole scaffold similar to other ARBs and is more potent and longer-lasting than the first clinically used ARB losartan (Junggren et al., 1996). While structures of several different GPCRs have been reported, the determination of a new GPCR structure remains a significant challenge. X-ray crystallography using synchrotron radiation requires sufficiently large crystals in order to collect high-resolution data. Our extensive efforts to solve the AT₁R structure were hampered by the limited size of micro-crystals grown in the membrane mimetic matrix known as lipidic cubic phase (LCP) (Caffrey and Cherezov, 2009). Nevertheless, by applying the recently developed method of serial femtosecond crystallography with LCP as a growth and carrier matrix for delivering microcrystals (LCP-SFX) into an X-ray free-electron laser (XFEL) beam (Liu et al., 2013; Weierstall et al., 2014; Liu et al., 2014a), we successfully determined the room-temperature crystal structure of the human AT₁R in complex with ZD7155 (AT₁R-ZD7155). Based on the AT₁R-ZD7155 structure, we further performed mutagenesis and docking simulations to reveal binding modes for clinically used anti-hypertensive drugs targeting AT₁R.

RESULTS

Structure Determination of AT₁R-ZD7155 Complex Using LCP-SFX Method

To facilitate crystallization, a thermostabilized apocytochrome, b₅₆₂RIL (BRIL) (Chun et al., 2012), was fused to the amino terminus (N terminus) of the human AT₁R. Eleven residues were trun-

cated from the N-terminal region of AT₁R (Met1, Thr7-Asp16), in order to shorten the flexible N terminus while keeping both the putative glycosylation site at Asn4 and the disulfide bond site at Cys18 intact. Forty residues were truncated from the carboxyl terminus (C terminus) after the cytoplasmic helix VIII (Figure 1A). The effect of protein engineering on AT₁R function was evaluated using radio-ligand binding and calcium mobilization assays, in which neither the truncations nor BRIL insertion significantly altered the functional and pharmacological properties of the receptor for ligand binding and signaling (Figure 1B–1D). With this engineered AT₁R, we obtained micro-crystals (maximum size 40 × 4 × 4 μm^3) in monoolein-based LCP, supplemented with cholesterol (Figure S1A). These microcrystals diffracted to only about 4-Å resolution at a synchrotron source under cryogenic conditions. To improve the resolution and avoid radiation damage and freezing, we took advantage of a recently developed LCP-SFX method and collected diffraction data at room temperature at the Linac Coherent Light Source (LCLS) using AT₁R micro-crystals (average size 10 × 2 × 2 μm^3) grown in syringes (Figures S1B and S1C). A total of 2,764,739 patterns were collected by using ~65 μl of crystal-loaded LCP, corresponding to ~0.35 mg of protein. Of these frames, 457,275 were identified as crystal hits, corresponding to a hit rate of 17%. Of these crystal hits, 73,130 frames (16%) were successfully indexed and integrated by CrystFEL (White et al., 2012) to 2.9-Å resolution (Table S1 and Figures S1D–S1F). The structure of the AT₁R-ZD7155 complex was refined to $R_{\text{work}}/R_{\text{free}}$ of 22.8%/27.4%. The final structure includes 289 out of 359 residues in the full-length human AT₁R (Figure 1A), and it has well-defined densities for most AT₁R residues and for the ligand ZD7155.

Overall Architecture of AT₁R

AT₁R, being the angiotensin II octapeptide receptor, shares some sequence similarity with other peptide receptors of class A GPCRs, structures of which are known (sequence alignment is shown in Figure S2), with the closest homology to the chemokine receptors (e.g., 36% sequence identity with CXCR4) and opioid receptors (e.g., 33% sequence identity with κ -OR) (Wu et al., 2010; Wu et al., 2012). AT₁R exhibits the canonical seven transmembrane α -helical (7TM) architecture, with an extracellular N terminus, three intracellular loops (ICL1-3), three extracellular loops (ECL1-3), an amphipathic helix VIII and an intracellular C terminus (Figure 2A). The overall fold of the angiotensin receptor AT₁R is most similar to the chemokine and opioid receptors (Figure 2B), with the lowest root mean square deviation for 80% of AT₁R α -carbon atoms (RMSD_{C α}) of about 1.8 Å to the nociceptin/orphanin FQ peptide receptor (NOP) (Thompson et al., 2012). Despite the overall similarity, a number of structural differences in the transmembrane bundle were observed between AT₁R and other peptide GPCRs (Figures 2C and 2D). For example, the tilts and extensions of the extracellular ends of helices I, V, VI, and VII are substantially different among these peptide receptors, while at the intracellular side, helices IV and V adopt the most diverse conformations. The conformations of helices II and III, however, are nearly identical for all these peptide receptors.

The extracellular part of AT₁R consists of the N-terminal segment, ECL1 (Glu91-Phe96) linking helices II and III, ECL2

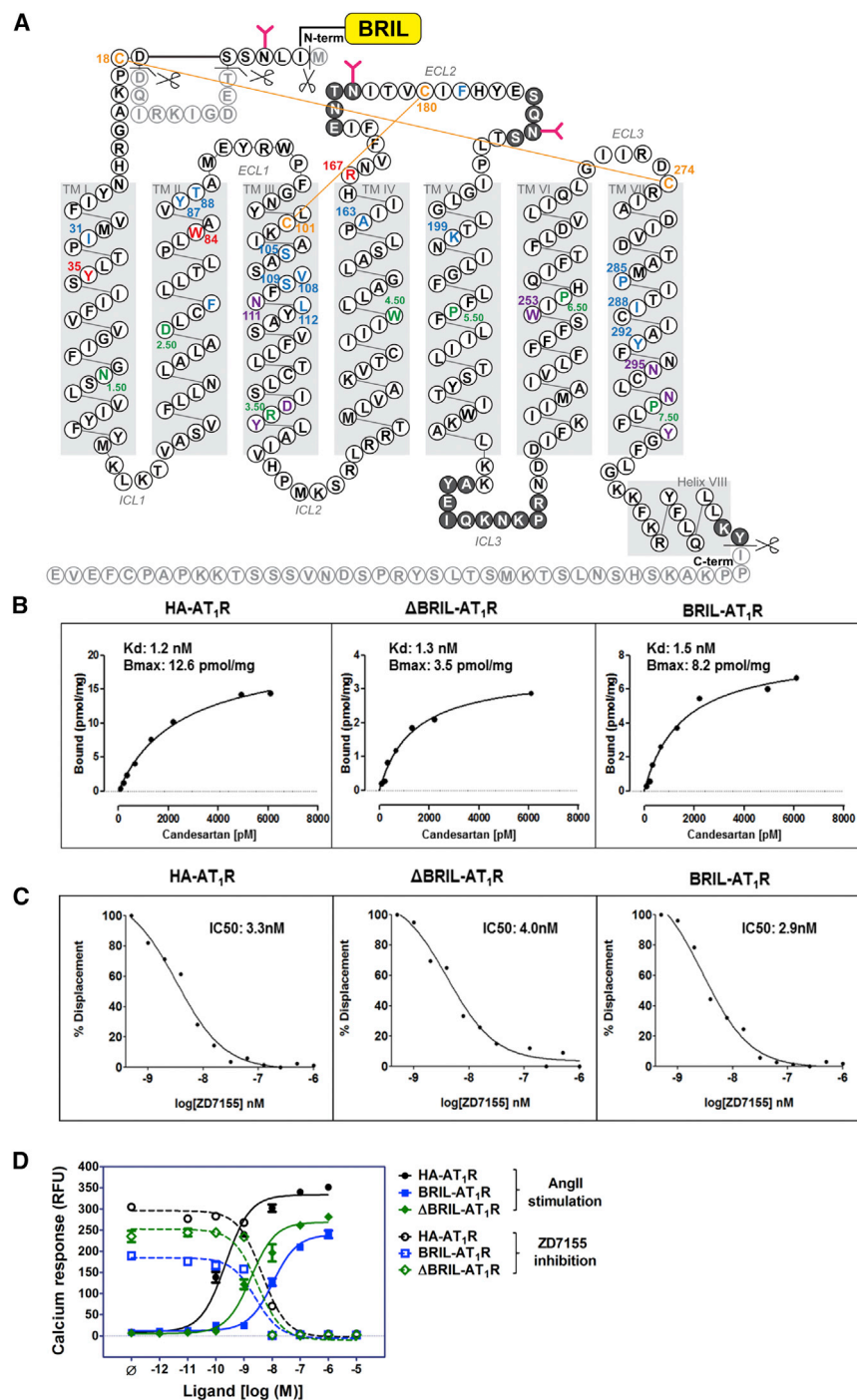


Figure 1. AT₁R Construct Design and Functional Characterization

(A) Snake plot of the BRIL-AT₁R construct used for crystallization. Residues that occupy the most conserved positions on each helix in class A GPCRs (X.50; B&W scheme) are colored in green. The four cysteine residues that form two disulfide bonds in the extracellular region are colored in orange. Three critical residues for ZD7155 binding are colored in red. All other residues that interact with ZD7155 are colored in blue. Critical residues/motifs for AT₁R activation are colored in purple. Truncated residues are shown as light gray, and residues that do not have sufficient density in the structure and therefore were not modeled are shown in dark gray circles.

(B) Saturation binding of the non-peptide antagonist ³H-candesartan to the wild-type HA-AT₁R, ΔBRIL-AT₁R, and BRIL-AT₁R.

(C) Competition binding of ZD7155 to the wild-type HA-AT₁R, ΔBRIL-AT₁R, and BRIL-AT₁R, performed by displacement of ³H-candesartan.

(D) Intracellular calcium responses for the wild-type HA-AT₁R, BRIL-AT₁R, and ΔBRIL-AT₁R. The agonist AngII and the antagonist ZD7155 dose-response curves for HA-AT₁R (circles), BRIL-AT₁R (squares), and ΔBRIL-AT₁R (diamonds) are shown in closed and open symbols, respectively. Error bars represent SEM.

mon motif among peptide GPCRs (Figure 2E). Intriguingly, ECL2 of AT₁R was found to serve as an epitope for the harmful agonistic autoantibodies in pre-eclampsia and malignant hypertension (Unal et al., 2012; Xia and Kellems, 2013).

The intracellular portion of AT₁R contains ICL1 (Lys58 to Val62) linking helices I and II, ICL2 (Val131 to Arg137) linking helices III and IV, ICL3 (Leu222 to Asn235) linking helices V and VI, and the C-terminal helix VIII. As in many other class A GPCRs, the conserved D(E)RY motif in helix III and the NPxxY motif in helix VII of AT₁R, both at the intracellular ends of transmembrane domain, were proposed to participate in receptor activation (Oliveira et al., 2007). However, the “ionic lock” salt bridge interaction between Arg^{3.50} (superscript indicates residue number as per the Ballesteros and Weinstein, 1995 [B&W] nomenclature) of the

(His166 to Ile191) linking helices IV and V, and ECL3 (Ile270 to Cys274) linking helices VI and VII (Figure 1A). Two disulfide bonds help to shape the extracellular side of AT₁R, with Cys18-Cys274 connecting the N terminus and ECL3, and Cys101-Cys180 connecting helix III and ECL2, similar to the chemokine receptors CXCR4 and CCR5 (Wu et al., 2010; Tan et al., 2013). Besides engaging in the conserved disulfide bonding, ECL2 of AT₁R exhibits a β-hairpin secondary structure, a com-

D(E)RY motif and Asp/Glu^{6.30} at the cytoplasmic end of helix VI is not possible in AT₁R, because the human AT₁R lacks an acidic residue at the position 6.30.

The C-terminal helix VIII of AT₁R was shown to bind the calcium-regulated effector protein, calmodulin (Thomas et al., 1999). Integrity of this region is also important for receptor internalization and coupling to G protein activation and signaling (Thomas et al., 1995; Sano et al., 1997). In most previously

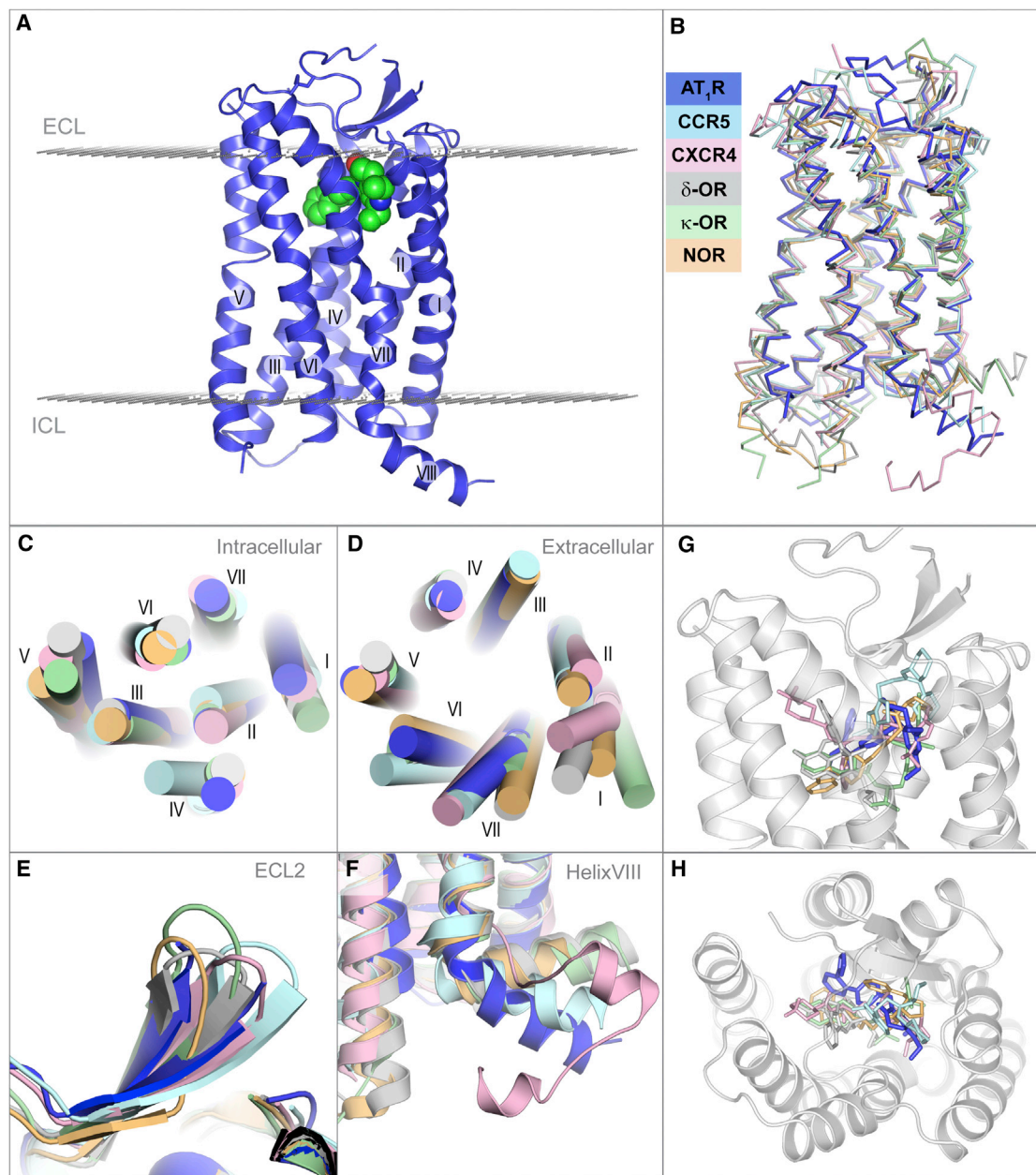


Figure 2. Overview of AT₁R-ZD7155 Architecture and Structural Comparison with Other Peptide GPCRs

(A) Overall AT₁R structure is shown as blue cartoon. ZD7155 is shown as spheres with carbon atoms colored green. Membrane boundaries, as defined by the PPM web server (Lomize et al., 2012), are shown as planes made of gray spheres.

(B–H) superposition of AT₁R with chemokine and opioid receptors, chemokine CCR5 receptor, light cyan (PDB ID 4MBS); chemokine CXCR4 receptor, light pink (PDB ID 3ODU); δ-opioid receptor, gray (PDB ID 4N6H); κ-opioid receptor, light green (PDB ID 4DJH); NOP receptor, light orange (PDB ID 4EA3), comparing the whole structure (B), intracellular view (C), extracellular view (D), ECL2 (E), helix VIII (F), and the ligand binding pocket side (G) and top (H) views.

See also Figures S1 and S2 and Table S1.

solved GPCR structures, helix VIII runs parallel to the membrane bilayer, however, in AT₁R it angles away from the membrane, resembling the orientation of this helix in CCR5 (Figure 2F). Experimentally, the secondary structure of AT₁R helix VIII was observed to be sensitive to hydrophobic environment, thereby associating with the cytoplasmic side of the cell membrane via

a high-affinity, anionic phospholipid-specific tethering that serves to increase the amphipathic helicity of this region (Mozzoli et al., 2002). As a separate peptide, helix VIII of AT₁R showed a higher affinity for lipid membranes that contained negatively charged phospholipids rather than zwitterionic phospholipids (Kamimori et al., 2005). A high concentration of positively

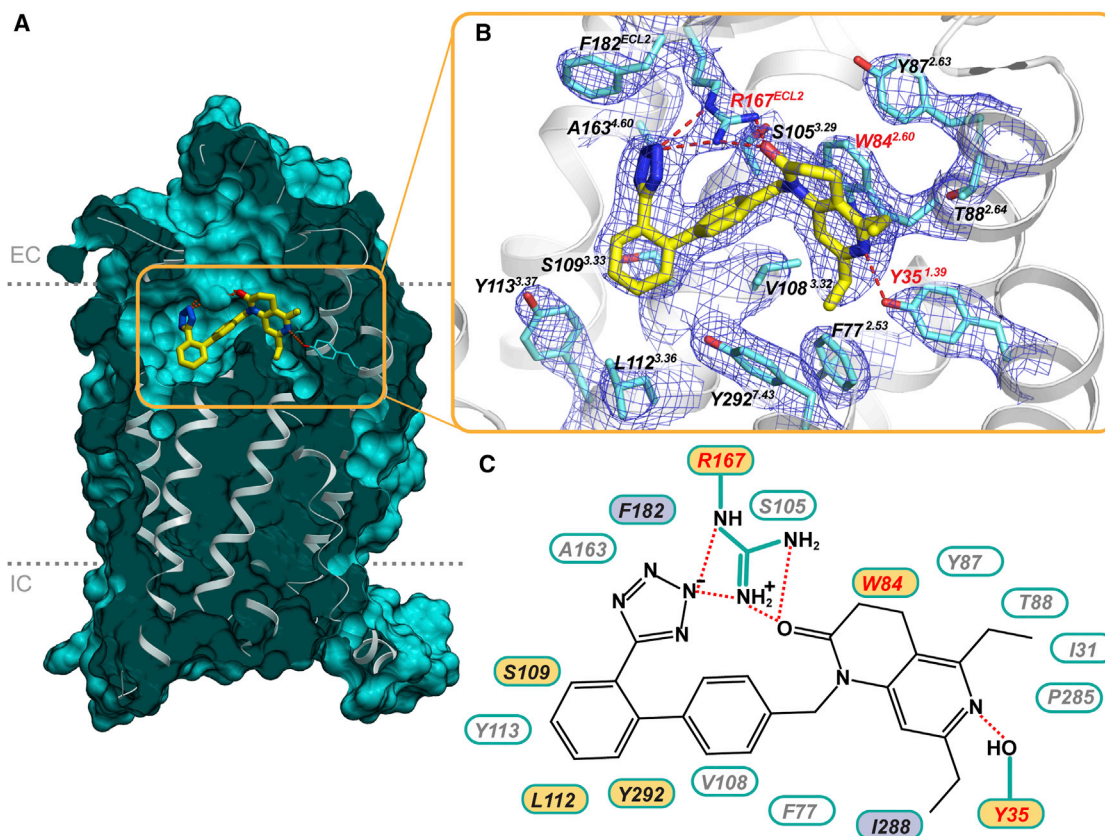


Figure 3. Interactions of ZD7155 with AT₁R

(A) Cross-section view of AT₁R highlighting the shape of the ligand binding pocket.

(B) Zoomed-in view of the ligand binding pocket showing all residues within 4 Å from the ligand ZD7155, along with the 2mFo-DFc electron density (blue mesh) contoured at 1 σ level. In (A) and (B) ZD7155 is shown as sticks with yellow carbons.

(C) Schematic representation of interactions between AT₁R and ZD7155. Hydrogen bonds/salt bridges are shown as red dashed lines. The residues shown by mutagenesis to be critical for ligand binding are labeled red, those that are important for either peptide or non-peptide ligands binding are labeled in yellow, and the residues that discriminate between peptide and non-peptide ligands are labeled in purple.

See also Figure S2 and Table S2.

charged residues (306-KKFKR-312) in helix VIII of AT₁R possibly defines its orientation and explains its sensitivity to the negatively charged lipids. Moreover, in AT₁R there is no putative palmitoylation site that is present in many GPCRs in this region, anchoring helix VIII to the lipid membrane.

ZD7155 Interactions in AT₁R Ligand-Binding Pocket

Small molecule antagonist ZD7155 was modeled into the prominent and well-defined electron density inside the ligand-binding pocket of AT₁R (Figure 3A and 3B), interacting with residues mainly from helices I, II, III, and VII, as well as ECL2. Side chains of Arg167^{ECL2} and Tyr35^{1.39} were found to form ionic and hydrogen bond interactions with ZD7155. The positively charged guanidine group of Arg167^{ECL2} forms an extensive interaction network with the acidic tetrazole and the naphthyridin-2-one moieties of ZD7155. Leveraging this information in mutagenesis studies, we found that mutation of Arg167^{ECL2} to alanine abolished both the peptide and non-peptide ligands binding to AT₁R (Table S2). However, the Arg167^{ECL2}Lys mutant showed only 2- to 3-fold reduced binding affinities for ZD7155, which

can be explained by the ability of lysine in this position to engage in salt bridge and hydrogen bond interactions similar to Arg167^{ECL2}, although likely with less optimal interaction geometry. The tetrazole moiety, or other acidic isostere in the ortho position of the biphenyl group comprises the most common scaffold among ARBs, and Arg167^{ECL2} is a unique residue of AT₁R compared to other structurally similar peptide GPCRs (Figure S2). This observation suggests that Arg167^{ECL2} may play an essential role in determining AT₁R ligand-binding affinity and selectivity. An additional hydrogen bond forms between Tyr35^{1.39} and the naphthyridin-2-one moiety of ZD7155. Our data showed that the Tyr35^{1.39}Ala mutant abolishes the binding capabilities of both peptide and non-peptide ligands with AT₁R (Table S2). Tyr^{1.39} is a well conserved residue in the angiotensin, chemokine, and opioid receptors (Figure S2). In the CCR5 structure, for example, Tyr37^{1.39} interacts with its ligand maraviroc (Tan et al., 2013).

The ZD7155 binding site in AT₁R partially overlaps with known ligand binding sites in the chemokine and opioid receptors (Figures 2G and 2H). Intriguingly, some of the residues that comprise

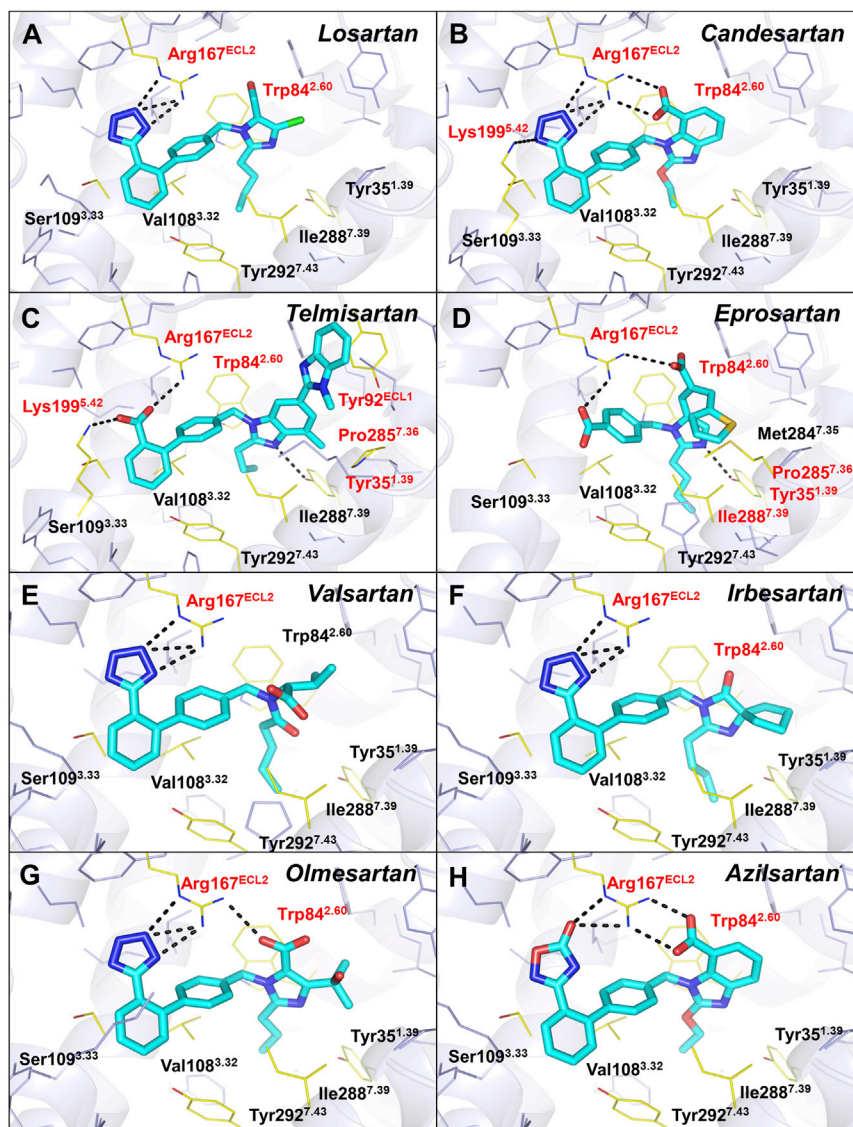


Figure 4. Docking of Different Anti-Hypertensive Drugs in the AT₁R Crystal Structure (A–H) The ARBs are shown as sticks with cyan carbons. The AT₁R residues interacting with ligands are labeled and shown as yellow lines, with the key residues highlighted in red. The hydrogen bonds are shown as black dashed lines. See also Table S3.

the other contacts for ZD7155 binding to AT₁R, however, are mediated by non-conserved residues, including Tyr87^{2.63}, Thr88^{2.64}, Ser105^{3.29}, Ser109^{3.33}, Ala163^{4.60}, Phe182^{ECL2}, Pro285^{7.36}, and Ile288^{7.39} (Figures 3B and 3C and Figure S2). These residues along with Arg167^{ECL2} therefore define the unique shape of the AT₁R ligand-binding pocket and explain the lack of cross-reactivity between ligands binding to AT₁R and other peptide receptors.

Binding Modes of Different ARBs toward AT₁R

To analyze the common and diverse features of the binding modes for different ARBs in AT₁R, we performed energy-based docking simulations of the clinically used anti-hypertensive ARBs using the AT₁R structure. The docking results show robust positioning of these compounds in the AT₁R ligand-binding pocket (Figure 4 and Table S3). Although the nature of the interactions with AT₁R is different for each ARB given their distinct chemical structures, most of these compounds are bound in similar orientations and engage in interactions with the three residues critical for ZD7155 binding, Arg167^{ECL2}, Trp84^{2.60}, and Tyr35^{1.39} (Figure 5). Residues Phe77^{2.53}, Tyr87^{2.63}, Ser105^{3.29}, Val108^{3.32}, Ser109^{3.33}, Leu112^{3.36}, Ala163^{4.60}, Phe182^{ECL2}, Ile288^{7.39}, and Tyr292^{7.43} also contribute to the receptor-ligand interactions and shape the ligand-binding pocket. For example, one of the common features among these ARBs is a short alkyl tail with two-four carbons extending into a narrow hydrophobic pocket formed by Tyr35^{1.39}, Phe77^{2.53}, Val108^{3.32}, Ile288^{7.39}, and Tyr292^{7.43} (Figure 5).

Losartan is the first clinically used ARB for the treatment of hypertension. It is, however, a surmountable antagonist with lower binding affinity to AT₁R compared to the later developed ARBs (Miura et al., 2011). Docking results suggest that Arg167^{ECL2} forms a salt bridge only with the tetrazole moiety of losartan but lacks polar interactions with other groups (Figure 4 and Table S3). Although the derived imidazole moiety of losartan can also contribute to polar interactions via methanol hydrogen bond to Cys180^{ECL2} main chain or via nitrogen interaction with

the ligand-binding pockets, including Ile^{1.35}, Phe^{2.53}, Trp^{2.60}, and Tyr^{7.43}, can be found among these structurally similar peptide GPCRs (Figure S2). Residues Phe77^{2.53} and Trp84^{2.60} from helix II of AT₁R are conserved in the chemokine receptors CXCR4 and CCR5 (Wu et al., 2010; Tan et al., 2013). Particularly, Trp84^{2.60} of AT₁R forms π - π interaction with the naphthyridin-2-one moiety of ZD7155, and mutation of Trp84^{2.60} to alanine abolished both the peptide and non-peptide ligands binding to AT₁R (Figure 3C and Table S2). Residues Ile31^{1.35} and Tyr292^{7.43} from helices I and VII of AT₁R are conserved in the opioid receptors κ -OR, δ -OR, and NOP. Additionally, residues Val108^{3.32} and Leu112^{3.36}, which hydrophobically interact with ZD7155 in the AT₁R ligand-binding pocket, are replaced by Trp108^{3.32} and Phe112^{3.36} in CCR5 and form hydrophobic interactions with its ligand maraviroc. In contrast, the position 3.32 in the aminergic and opioid receptors is occupied by a conserved aspartic acid that engages in a salt bridge interaction with ligands. Most of

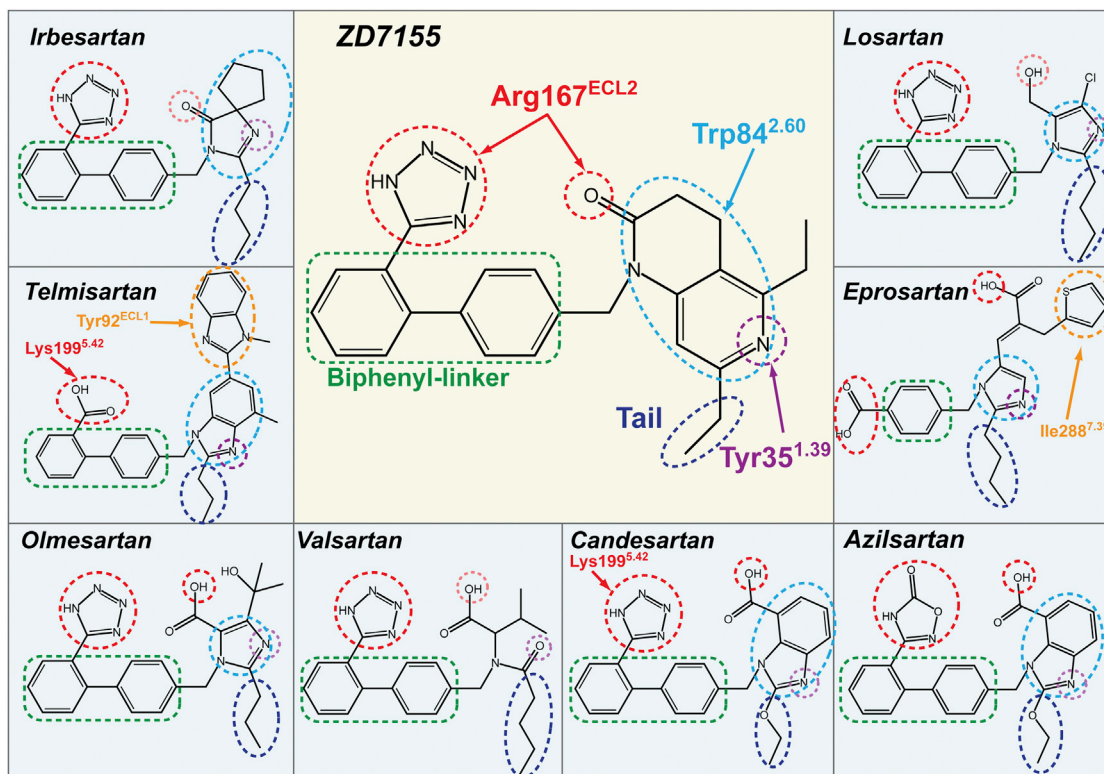


Figure 5. Common and Distinct Binding Modes of Different ARBs with AT₁R

The ARB chemical groups that are engaged in hydrogen bonding/salt bridging with Arg167^{ECL2} and Tyr35^{1.39} are marked by red and purple dashed circles, respectively. Pale red and pale purple dotted circles are used for groups with sub-optimal contacts as suggested by docking. The heterocyclic groups forming π - π contacts with Trp84^{2.60} are surrounded by light-blue dashed circles. The biphenyl-linker groups for hydrophobic interactions are outlined by green dashed boxes, and the two-four carbons tails, extending into the hydrophobic pocket formed by Tyr35^{1.39}, Phe77^{2.53}, Val108^{3.32}, Ile288^{7.39}, and Tyr292^{7.43}, are outlined by dark-blue dashed circles. Specific interactions of candesartan and telmisartan with Lys199^{5.42} are shown by red arrows. Specific interactions between Tyr92^{ECL1} and telmisartan, and between Ile288^{7.39} and eprosartan are highlighted by orange dashed circles.

See also Figure S3.

Tyr35^{1.39}, distances and angles for hydrogen bonding are sub-optimal; this may explain the lower binding affinity and surmountable property of losartan at AT₁R. An active metabolite of losartan, EXP3174, is predicted to bind in a similar pose as losartan, but instead of interaction with Cys180^{ECL2}, its carboxyl group could engage in a second salt bridge interaction with Arg167^{ECL2}, similarly to ZD7155 (Table S3). In contrast, candesartan is an insurmountable inverse agonist with a slow dissociation rate from AT₁R (Takezako et al., 2004). The docking results indicate that besides interacting with the tetrazole moiety of candesartan, Arg167^{ECL2} forms two salt bridges to the carboxylic group of the benzimidazole moiety (Figure 4 and Table S3). Moreover, Lys199^{5.42} is predicted to form an additional salt bridge with the tetrazole moiety, which can further stabilize candesartan binding. Telmisartan lacks the conserved tetrazole moiety among ARBs. Instead, the carboxylic group of telmisartan is predicted to form salt bridges with both Arg167^{ECL2} and Lys199^{5.42} (Figure 4 and Table S3). Moreover, unlike other ARBs studied here, two consecutive benzimidazole moieties of telmisartan extend to Tyr92^{ECL1}, making additional hydrophobic and π - π contacts, which are likely to contribute to its high potency (Balakumar et al., 2012). This prediction was confirmed

by our mutagenesis data, which showed a dramatic decrease in affinity of telmisartan to the Tyr92^{ECL1} Ala mutant (Figure S3A). Eprosartan is the most unique among the ARBs studied here, lacking both the tetrazole group and one of the two benzene rings of the biphenyl scaffold. As our docking results suggest, eprosartan uses its two carboxyl groups to form salt bridges with Arg167^{ECL2} (Figure 4 and Table S3). Additionally, the specific thiophen moiety of eprosartan forms hydrophobic interactions with Pro285^{7.36} and Ile288^{7.39} and reaches toward Met284^{7.35}. Mutation of Met284^{7.35} to alanine produced minimal effect, slightly increasing the affinity for eprosartan binding, in agreement with predicted interactions of this ligand with only mainchain and C β atoms of Met284^{7.35} (Figure S3B). On the other hand, mutations Pro285^{7.36}Ala and Ile288^{7.39}Ala induced a strong decrease in the binding affinity of eprosartan (Figures S3C and S3D), highlighting essential role of these residues in eprosartan binding. Finally, both our crystal structure and docking results suggest that Lys199^{5.42} retains some conformational heterogeneity in AT₁R. Docking with the flexible side chain of Lys199^{5.42} indicates that the amino group of this residue can reach the acidic moieties of ARBs by forming salt bridges (as interacting with candesartan and telmisartan) or water-mediated

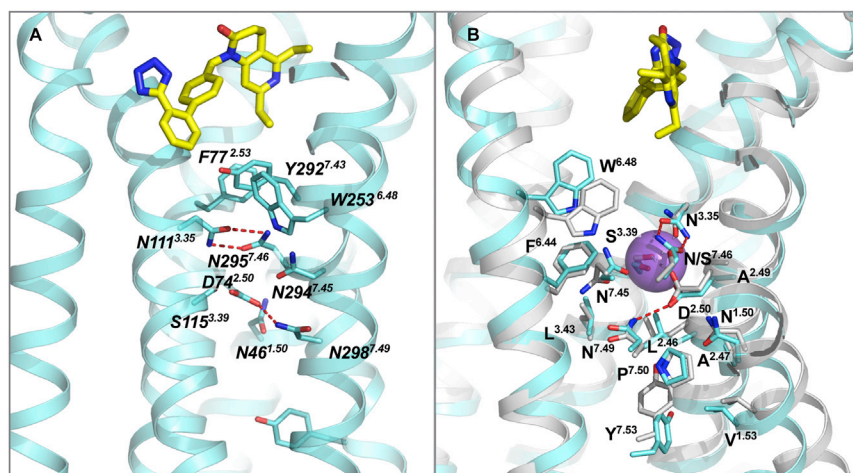


Figure 6. Critical Residues for AT₁R Activation

(A) A cluster of aromatic residues (F77^{2.53}, W253^{6.48} and Y292^{7.43}) is located just below ZD7155, bridging the ligand binding pocket with a cluster of polar residues that includes several highly conserved in class A GPCR residues (N46^{1.50}, D74^{2.50}), along with N111^{3.35} and N295^{7.46} forming hydrogen bonds that hold helices III and VII together.

(B) Superposition of the AT₁R structure with the high-resolution structure of δ -OR (PDB ID 4N6H) reveals a high structural conservation of the putative sodium-binding site. Sodium ion is shown as purple ball.

interactions, which may explain the reduced ligand-binding capabilities of Lys199^{5.42} mutants (Table S2).

Mechanism of AT₁R Modulation

Based on previous observations that mutations of either Asn111^{3.35} or Asn295^{7.46} induce constitutive activation of the receptor, it was proposed that the inactive conformation of AT₁R is stabilized by interactions between Asn111^{3.35} and Asn295^{7.46}. Further, it was suggested that binding of AngII to the wild-type (WT) receptor disrupts the hydrogen bonds between Asn111^{3.35} and Asn295^{7.46}, thus allowing Asn295^{7.46} to interact with the conserved Asp74^{2.50} (Balakumar and Jagadeesh, 2014; Unal and Karnik, 2014). Indeed, two intramolecular hydrogen bonds are observed between Asn111^{3.35} and Asn295^{7.46} in the AT₁R-ZD7155 structure (Figure 6A). Of particular interest, Asp74^{2.50}, Asn111^{3.35}, and Asn295^{7.46}, together with two other residues, Trp253^{6.48} from the WxP motif and Asn298^{7.49} from the NPxxY motif, belong to the putative sodium pocket of AT₁R (Katritch et al., 2014) as revealed by superposition with the sodium site in the high-resolution structure of δ -OR (Figure 6B) (Fenalti et al., 2014). All residues lining this pocket in AT₁R are conserved exactly as in δ -OR, except for Asn295^{7.46} (Ser in δ -OR), which is observed at this position in a GPCR structure for the first time; therefore, its presence and the strong hydrogen bond interactions with Asn111^{3.35} may impact the sodium binding and functional properties of AT₁R. Moreover, the neighboring residue Phe77^{2.53} from the ligand-binding pocket of AT₁R was also found to be critical for the inter-helical interactions required for AT₁R activation (Miura et al., 2003). Combination of Phe77^{2.53}Ala and Asn111^{3.35}Gly mutations resulted in an almost fully active receptor (Miura et al., 2008). Thus, multiple structural and functional data suggest that the hydrogen bond network around Asn111^{3.35} and Asn295^{7.46} as revealed in the current structure may play an essential role in AT₁R activation, probably by relaying the conformational changes in the ligand-binding pocket to the cytoplasmic domain coupling to the downstream signaling, although further structural, functional, and biophysical studies are required to fully understand the mechanism of AT₁R modulation.

DISCUSSION

The angiotensin receptor AT₁R is a therapeutic target of outstanding interest due to its important roles in cardiovascular pathophysiology. Several AT₁R blockers have been developed and clinically used as anti-hypertensive drugs. Although extensive efforts were taken to delineate the pharmacophores of AT₁R ligands, structure-based drug design was still hindered by the lack of structural information. By using an XFEL, we successfully determined the crystal structure of the human AT₁R in complex with its antagonist ZD7155. Compared to the traditional X-ray crystallography with cryo-cooled crystals, the LCP-SFX method yields the room-temperature structure of the AT₁R-ZD7155 complex, which is likely to represent more accurately the receptor conformations and dynamics in the native cellular environment. The AT₁R-ZD7155 complex structure reveals a variety of key features of AT₁R shared with other GPCR family members, as well as many novel and unique structural characteristics of the angiotensin receptor. Unexpectedly, three AT₁R residues, which have not been previously implicated in binding small molecule ligands, were found to form critical interactions with ZD7155; Arg167^{ECL2} and Tyr35^{1.39} are engaged in ionic and hydrogen bonds, while Trp84^{2.60} forms extensive π - π interactions with the ligand. The antagonist-bound AT₁R structure was used further for docking of several anti-hypertensive ARBs into the AT₁R ligand-binding pocket, elucidating the structural basis for AT₁R modulation by drugs. Our extensive mutagenesis experiments revealed that residues Tyr35^{1.39}, Trp84^{2.60}, Arg167^{ECL2}, and Lys199^{5.42} are critical for both peptide ([Sar¹, Ile⁸]-AngII) and non-peptide (candesartan) binding. Residues Phe182^{ECL2} and Ile288^{7.39} discriminate between the peptide and non-peptide ligand (these mutants do not bind [Sar¹, Ile⁸]-AngII but bind candesartan). Mutations of Ser109^{3.33} and Tyr292^{7.43} slightly affected non-peptide (candesartan) binding but not peptide binding (Table S2).

Among the naturally occurring amino acid variations in AT₁R, reported in Uniprot (<http://www.uniprot.org/uniprot/P30556>), Ala163^{4.60}Thr, Thr282^{7.33}Met, and Cys289^{7.40}Trp are located near the binding pocket for ARBs. These variants may directly

alter binding of ARBs and therefore modify the anti-hypertensive response to treatment with different ARBs in individuals carrying these variations. In contrast, Leu48^{1,52}Val, Leu222^{ICL3}Val, and Ala244^{6,39}Ser, which are located closer to intracellular ends of helices, may indirectly influence binding of ARBs or signaling by AT₁R. Finally, Thr336Pro and Pro341His are located in the C-terminal tail that was not included in the crystallized construct. These residues, however, are known to affect GPCR kinase-dependent phosphorylation, an event that is necessary for β -arrestin recruitment to AT₁R.

Of particular interest, the atomic details of ECL2 and the extracellular ligand-binding region, revealed in the current structure, are expected to guide design of two different types of therapeutic agents targeting AT₁R, the anti-hypertensive ARBs extensively interacting with Arg167^{ECL2} on the ligand-binding pocket side of ECL2, and the peptide-mimicking antigens against auto-antibodies, which bind to the extracellular side of ECL2 in patients with autoimmune disorders, such as preeclampsia and malignant hypertension (Zhou et al., 2008; Fu et al., 2000). Therefore, our results provide long anticipated insights into the AT₁R structure-function relationship and pharmacological properties and demonstrate the potential for using the LCP-SFX method at XFEL sources to accelerate structural studies of challenging targets.

EXPERIMENTAL PROCEDURES

Protein Engineering for Structural Studies

The sequence of the human AT₁R gene was optimized for insect cell expression and synthesized by GenScript. A thermostabilized apocytochrome *b*₅₆₂RIL (BRIL) from *E. coli* (M7W, H102I, R106L) was fused to the N terminus of the human AT₁R, using overlapping PCR. The construct has truncations of the AT₁R residues 1, 7–16, and 320–359. The resulting BRIL-AT₁R chimera sequence was subcloned into a modified pFastBac1 vector (Invitrogen), which contained a haemagglutinin (HA) signal sequence, a FLAG tag and 10 \times His tag, followed by a tobacco etch virus (TEV) protease cleavage site, before the N terminus of the chimera sequence.

Protein Expression and Purification

BRIL-AT₁R construct was expressed in *Spodoptera frugiperda* (Sf9) insect cells using the Bac-to-Bac baculovirus expression system (Invitrogen). Cells with a density of $2\text{--}3 \times 10^6$ cells per ml were infected with baculovirus at 27°C, and harvested at 48 hr after infection.

BRIL-AT₁R in complex with ZD7155 (Tocris Bioscience) was solubilized from isolated membranes using 1% (w/v) n-dodecyl-beta-D-maltopyranoside (DDM, Anatrace) and 0.2% (w/v) cholesterol hemisuccinate (CHS, Sigma-Aldrich). After purification by metal affinity chromatography BRIL-AT₁R/ZD7155 complex was desalted to remove imidazole using PD MiniTrap G-25 column (GE Healthcare) and then treated overnight with His-tagged TEV protease to cleave the N-terminal FLAG/His tags from the protein. The cleaved FLAG/His tags and TEV protease were removed by TALON IMAC resin. The protein was not treated with PNGase F and therefore remained fully glycosylated. Finally, the purified protein was concentrated to 30 mg/ml with a 100 kDa cutoff concentrator (Vivaspin) and used in crystallization trials. The protein yield and monodispersity were tested by analytical size exclusion chromatography (aSEC).

Lipidic Cubic Phase Crystallization

BRIL-AT₁R in complex with ZD7155 was crystallized in LCP composed of monoolein supplemented with 10% cholesterol (Caffrey and Cherezov, 2009). LCP crystallization trials were performed using an NT8-LCP crystallization robot (Formulatrix). 96-well glass sandwich plates (Marienfeld) were incubated and imaged at 20°C using an automatic incubator/imager (RockImager

1000, Formulatrix). The crystals grew in the condition of 100 mM sodium citrate (pH 5.0–6.0), 300–600 mM NH₄H₂PO₄, 20%–30% (v/v) PEG400 and 2%–8% (v/v) DMSO. The crystals were harvested using micromounts (MiTeGen) and flash-frozen in liquid nitrogen for data collection at a synchrotron source. These crystals diffracted only to about 4-Å resolution, even after extensive optimization of crystallization conditions.

Microcrystals for SFX data collection were prepared in gas-tight syringes (Hamilton) as described (Liu et al., 2014b), using 100 mM sodium citrate (pH 5.0), 450 mM NH₄H₂PO₄, 28% (v/v) PEG400 and 4% (v/v) DMSO as a precipitant. Before loading microcrystals in the LCP injector the excess precipitant was removed, and 7.9 MAG was added and mixed with LCP, to absorb the residual precipitant solution and prevent formation of a crystalline phase due to a rapid evaporative cooling when injecting LCP into vacuum (Weierstall et al., 2014).

X-Ray Free Electron Laser Data Collection

Data collection was performed at the Coherent X-ray Imaging (CXI) end station of the Linac Coherent Light Source (LCLS), SLAC National Accelerator Laboratory, using XFEL pulses of 36 fs duration focused to a size of $1.5 \times 1.5 \mu\text{m}^2$ by Kirkpatrick-Baez mirrors. A photon energy of 7.9 keV, an average pulse energy of 2.7 mJ and a transmission level of 16% resulted in a maximum dose of 75 MGy at the sample.

Microcrystals dispersed in LCP were delivered into the interaction region using an LCP injector (Weierstall et al., 2014) with a 50 μm diameter nozzle at a flow rate of 170 nl per minute. Diffraction patterns were collected on a Cornell-SLAC Pixel array detector (CSPAD - version 1.5) (Hart et al., 2012) at a rate of 120 Hz.

With a total sample volume of 65 μl , a total of 2,764,739 diffraction frames were collected within 6.4 hr. Initial frames were corrected and filtered using the software package Cheetah (Barty et al., 2014). A crystal “hit” was defined as an image containing a minimum of 15 diffraction peaks with a signal to noise ratio above 4. A total of 457,275 positive “hits” were further processed using the CrystFEL software suite (version 0.5.3) (White et al., 2012). The detector geometry was refined using an automated algorithm designed to match found and predicted peaks to sub-pixel accuracy. By further refinement of parameters (peak detection, prediction, and integration), a total of 73,130 images were indexed, integrated, and merged into a final dataset. To reduce noise and outliers and thus improve data quality we have applied two data rejection criteria: (1) per pattern resolution cutoff, and (2) rejection of patterns based on a Pearson correlation coefficient threshold, as described in the [Extended Experimental Procedures](#). A resolution cutoff was estimated to be 2.9 Å using a combination of CC* (Karplus and Diederichs, 2012) and other parameters (Figures S1D–S1F). The final dataset had overall $R_{\text{split}} = 9.8\%$, and CC* = 0.872 in the highest resolution shell.

Structure Determination

The structure was solved by molecular replacement with Phaser (McCoy et al., 2007) using an automated script described in the [Extended Experimental Procedures](#).

Refinement and model completion were performed by repetitive cycling between Refmac5 (Murshudov et al., 1997) and autoBUSTER (Bricogne et al., 2009), followed by manual examination and rebuilding of the refined coordinates in Coot (Emsley et al., 2010). Data collection and refinement statistics are shown in Table S1.

Docking of ARBs into AT₁R Ligand-Binding Pocket

Representative ARBs were docked into the AT₁R crystal structure using an energy-based docking protocol implemented in ICM molecular modeling software suite (Molsoft). Molecular models of compounds were generated from two-dimensional representations and their 3D geometry was optimized using MMFF-94 force field (Halgren, 1995). Molecular docking employed biased probability Monte Carlo (BPMC) optimization of the ligand internal coordinates in the grid potentials of the receptor (Totrov and Abagyan, 1997). To ensure convergence of the docking procedure, at least five independent docking runs were performed for each ligand starting from a random conformation. The results of individual docking runs for each ligand were considered consistent if at least three of the five docking runs produced similar ligand

conformations (RMSD < 2.0 Å) and Binding Score < −20.0 kJ/mol. The unbiased docking procedure did not use distance restraints or any other a priori derived information for the ligand-receptor interactions.

Ligand Binding Assays

Ligand binding was analyzed using total membranes prepared from COS-1 cells transiently expressing HA-AT₁R (wild-type), ΔBRIL-AT₁R (crystallized construct without BRIL), and BRIL-AT₁R (crystallized construct) constructs. Single mutants were constructed by a PCR-based site-directed mutagenesis strategy as previously described (Unal et al., 2010). Protein concentration was determined by Bio-Rad Protein Assay (Bio-Rad). For both saturation and competition binding assays, 10 μg of homogenous cell membrane was used per well.

Saturation binding assays with ³H-candesartan were performed under equilibrium conditions, with ³H-candesartan (Amersham Pharmacia Biotech) concentrations ranging between 0.125 and 12 nM (specific activity, 16 Ci/mmol) as duplicates in 96-well plates for 1 hr at room temperature. Nonspecific binding was measured in the presence of 10 μM candesartan (gift from AstraZeneca). The binding kinetics were analyzed by nonlinear curve-fitting program GraphPad Prism 5, which yielded the mean ± SD for the K_d and B_{max} values.

Competition binding assays were performed under equilibrium conditions, with 2 nM ³H-candesartan and various concentrations of the ZD7155 ranging between 0.04 and 1,000 nM. The binding kinetics were analyzed by nonlinear curve-fitting program GraphPad Prism 5, which yielded the mean ± SD for the IC₅₀ values.

Signaling Assays in Whole Cells

Calcium levels inside COS-1 cells transiently expressing different AT₁R constructs were measured using a Fluorescent Imaging Plate Reader (FLIPR) Calcium 5 assay kit (Molecular Devices). For the antagonist dose-response, the cells were first treated with different concentrations of ZD7155 for 1 hr followed by stimulation with 100 nM AngII. The EC₅₀ values for AngII dose response were 0.2, 2, and 12 nM for HA-AT₁R, ΔBRIL-AT₁R, and BRIL-AT₁R, respectively. The IC₅₀ values for ZD7155 to inhibit AngII response were between 3 to 4 nM for all constructs. The curves from a representative experiment wherein measurements are made in triplicate are shown as mean ± SEM. Additional information is available in the [Extended Experimental Procedures](#).

ACCESSION NUMBERS

The coordinates and structure factors have been deposited into the Protein Data Bank under the accession code 4YAY.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, three figures, and three tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.04.011>.

AUTHOR CONTRIBUTIONS

H.Z. designed, optimized, purified, and characterized receptor constructs for structural studies, crystallized the receptor in LCP; collected and processed diffraction data; determined the structure; analyzed the data; and wrote the paper. H.U. performed mutagenesis, signaling, and ligand binding studies and contributed to writing the paper. C.G. participated in the XFEL data collection and processed XFEL data. G.W.H. solved and refined the AT₁R structure. N.A.Z. participated in XFEL data collection and contributed in the XFEL data processing. D.J., D.W., G.N., and U.W. designed, prepared, and operated the LCP injector during the XFEL data collection. M.M., G.J.W., and S.B. operated the CXI beamline and performed the XFEL data collection. O.M.Y. refined the geometry of the CSPAD detector. T.A.W. implemented new data processing algorithms in CrystFEL, used in this study. W.L., C.W. and A.I. helped with XFEL sample preparation and participated in the XFEL data collection. K.C.T. and R.D. participated in mutagenesis, membrane production, signaling, and ELISA data collection and analysis. M.R.S. and Q.X. helped with synchrotron

data processing, XFEL structure solution and structure validation. J.C., C.E.C. and P.F. helped with biophysical characterization of microcrystals at LCLS and participated in the XFEL data collection. R.C.S. conceived the project, supervised receptor expression and characterization, and contributed to writing the paper. V.K. designed initial AT₁R constructs, analyzed the structure, performed docking studies, and wrote the paper. S.S.K. conceived the project, supervised mutagenesis and functional studies, and contributed to writing the paper. V.C. conceived the project, supervised crystallization and crystallographic data collection, analyzed the data and wrote the paper. Edits were provided by C.G., G.W.H., T.A.W., U.W. and P.F.

ACKNOWLEDGMENTS

This work was supported in parts by the NIH grants R01 GM108635 (V.C.), U54 GM094618 (target GPCR-11) (V.K., V.C., R.C.S.), P41 GM103393 (S.B.), R01 HL57470 (S.S.K.), R01 HL115964 (S.S.K.), U54 GM094599 (P.F.), R01 GM095583 (P.F.), U54 GM094586 (Q.X.), and National Research Service award HL007914 (H.U.). Further support was provided by the NSF BioXFEL Science and Technology center grant 1231306 (P.F., U.W., G.N.); the Helmholtz Gemeinschaft, the DFG Cluster of Excellence "Center for Ultrafast Imaging"; the BMBF project FKZ 05K12CH1 (C.G., O.M.Y., T.A.W.); the PIER Helmholtz Graduate School and the Helmholtz Association (C.G.); and the Chinese 1000 Talent Program (R.C.S.). Parts of this research were carried out at the LCLS, a National User Facility operated by Stanford University on behalf of the U.S. Department of Energy, Office of Basic Energy Sciences, and at the GM/CA CAT of the Argonne Photon Source, Argonne National Laboratory. We thank J. Velasquez for help with molecular biology; T. Trinh and M. Chu for help with baculovirus expression; M. Metz, D. Oberthuer, A. Barty, S. Basu, and R. Fromme for help with the XFEL data collection and analysis, and R. Miller, K. Kadyshkevskaya, and A. Walker for assistance with manuscript preparation. We are grateful to J. Spence and H. Chapman for their encouragement and support.

Received: November 6, 2014

Revised: December 30, 2014

Accepted: March 2, 2015

Published: April 23, 2015

REFERENCES

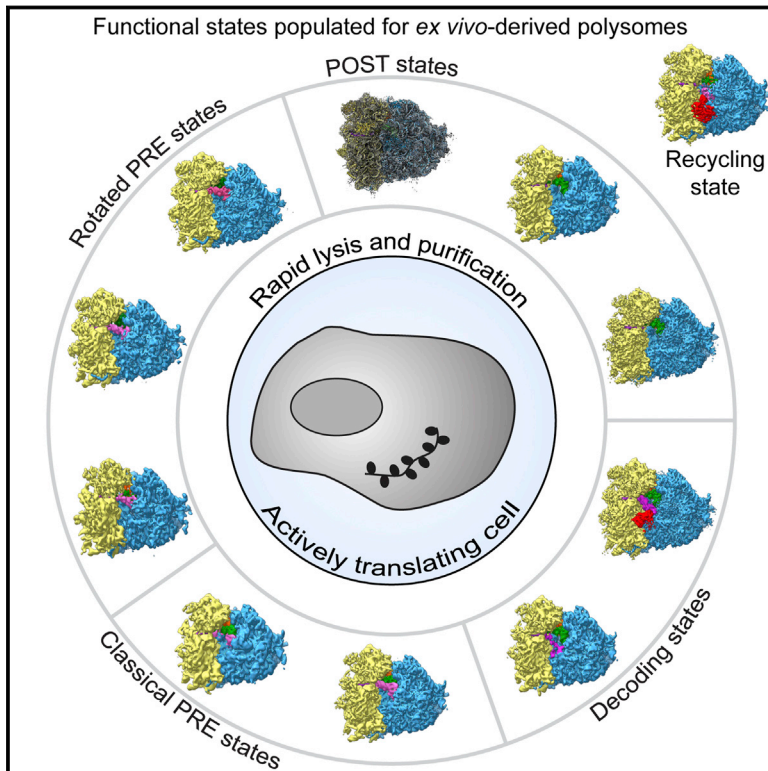
- Akazawa, H., Yano, M., Yabumoto, C., Kudo-Sakamoto, Y., and Komuro, I. (2013). Angiotensin II type 1 and type 2 receptor-induced cell signaling. *Curr. Pharm. Des.* 19, 2988–2995.
- Balakumar, P., and Jagadeesh, G. (2014). Structural determinants for binding, activation, and functional selectivity of the angiotensin AT1 receptor. *J. Mol. Endocrinol.* 53, R71–R92.
- Balakumar, P., Bishnoi, H.K., and Mahadevan, N. (2012). Telmisartan in the management of diabetic nephropathy: a contemporary view. *Curr. Diabetes Rev.* 8, 183–190.
- Ballesteros, J.A., and Weinstein, H. (1995). Integrated methods for the construction of three dimensional models and computational probing of structure-function relations in G-protein coupled receptors. *Methods Neurosci* 25, 366–428.
- Barty, A., Kirian, R.A., Maia, F.R., Hantke, M., Yoon, C.H., White, T.A., and Chapman, H. (2014). Cheetah: software for high-throughput reduction and analysis of serial femtosecond X-ray diffraction data. *J. Appl. Cryst.* 47, 1118–1131.
- Bricogne, G., Blanc, E., Brandl, M., Flensburg, C., Keller, P., Paciorek, W., Roversi, P., Sharff, A., Smart, O.S., Vonnrhein, C., et al. (2009). Buster version 2.8.0 (Cambridge, United Kingdom: Global Phasing Ltd).
- Burnier, M., and Brunner, H.R. (2000). Angiotensin II receptor antagonists. *Lancet* 355, 637–645.
- Caffrey, M., and Cherezov, V. (2009). Crystallizing membrane proteins using lipidic mesophases. *Nat. Protoc.* 4, 706–731.

- Chun, E., Thompson, A.A., Liu, W., Roth, C.B., Griffith, M.T., Katritch, V., Kunken, J., Xu, F., Cherezov, V., Hanson, M.A., and Stevens, R.C. (2012). Fusion partner toolchest for the stabilization and crystallization of G protein-coupled receptors. *Structure* 20, 967–976.
- de Gasparo, M., Catt, K.J., Inagami, T., Wright, J.W., and Unger, T. (2000). International union of pharmacology. XXIII. The angiotensin II receptors. *Pharmacol. Rev.* 52, 415–472.
- Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* 66, 486–501.
- Fenalti, G., Giguere, P.M., Katritch, V., Huang, X.P., Thompson, A.A., Cherezov, V., Roth, B.L., and Stevens, R.C. (2014). Molecular control of δ -opioid receptor signalling. *Nature* 506, 191–196.
- Fu, M.L., Herlitz, H., Schulze, W., Wallukat, G., Micke, P., Eftekhari, P., Sjögren, K.G., Hjalmarson, A., Müller-Esterl, W., and Hoebeke, J. (2000). Autoantibodies against the angiotensin receptor (AT1) in patients with hypertension. *J. Hypertens.* 18, 945–953.
- Halgren, T. (1995). Merck molecular force field I-V. *J. Comp Chem* 17, 490–641.
- Hart, P., Boutet, S., Carini, G., Dragone, A., Duda, B., Freytag, D., Haller, G., Herbst, R., Herrmann, S., Kenney, C., et al. (2012). The Cornell-SLAC Pixel Array Detector at LCLS. In *Nuclear Science Symposium, Medical Imaging Conference*.
- Imaizumi, S., Miura, S., Yahiro, E., Uehara, Y., Komuro, I., and Saku, K. (2013). Class- and molecule-specific differential effects of angiotensin II type 1 receptor blockers. *Curr. Pharm. Des.* 19, 3002–3008.
- Junggren, I.L., Zhao, X., Sun, X., and Hedner, T. (1996). Comparative cardiovascular effects of the angiotensin II type 1 receptor antagonists ZD 7155 and losartan in the rat. *J. Pharm. Pharmacol.* 48, 829–833.
- Kamimori, H., Unabia, S., Thomas, W.G., and Aguilar, M.I. (2005). Evaluation of the membrane-binding properties of the proximal region of the angiotensin II receptor (AT1A) carboxyl terminus by surface plasmon resonance. *Anal. Sci.* 21, 171–174.
- Karplus, P.A., and Diederichs, K. (2012). Linking crystallographic model and data quality. *Science* 336, 1030–1033.
- Katritch, V., Fenalti, G., Abola, E.E., Roth, B.L., Cherezov, V., and Stevens, R.C. (2014). Allosteric sodium in class A GPCR signaling. *Trends Biochem. Sci.* 39, 233–244.
- Keane, M.G., and Pyeritz, R.E. (2008). Medical management of Marfan syndrome. *Circulation* 117, 2802–2813.
- Liu, W., Wacker, D., Gati, C., Han, G.W., James, D., Wang, D., Nelson, G., Weierstall, U., Katritch, V., Barty, A., et al. (2013). Serial femtosecond crystallography of G protein-coupled receptors. *Science* 342, 1521–1524.
- Liu, W., Wacker, D., Wang, C., Abola, E., and Cherezov, V. (2014a). Femtosecond crystallography of membrane proteins in the lipidic cubic phase. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 369, 20130314.
- Liu, W., Ishchenko, A., and Cherezov, V. (2014b). Preparation of microcrystals in lipidic cubic phase for serial femtosecond crystallography. *Nat. Protoc.* 9, 2123–2134.
- Lomize, M.A., Pogozheva, I.D., Joo, H., Mosberg, H.I., and Lomize, A.L. (2012). OPM database and PPM web server: resources for positioning of proteins in membranes. *Nucleic Acids Res.* 40, D370–D376.
- Ma, T.K., Kam, K.K., Yan, B.P., and Lam, Y.Y. (2010). Renin-angiotensin-aldosterone system blockade for cardiovascular diseases: current status. *Br. J. Pharmacol.* 160, 1273–1292.
- McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. *J. Appl. Cryst.* 40, 658–674.
- Miura, S., Zhang, J., Boros, J., and Karnik, S.S. (2003). TM2-TM7 interaction in coupling movement of transmembrane helices to activation of the angiotensin II type-1 receptor. *J. Biol. Chem.* 278, 3720–3725.
- Miura, S., Kiya, Y., Kanazawa, T., Imaizumi, S., Fujino, M., Matsuo, Y., Karnik, S.S., and Saku, K. (2008). Differential bonding interactions of inverse agonists of angiotensin II type 1 receptor in stabilizing the inactive state. *Mol. Endocrinol.* 22, 139–146.
- Miura, S., Karnik, S.S., and Saku, K. (2011). Review: angiotensin II type 1 receptor blockers: class effects versus molecular effects. *J. Renin Angiotensin Aldosterone Syst.* 12, 1–7.
- Miura, S., Nakao, N., Hanzawa, H., Matsuo, Y., Saku, K., and Karnik, S.S. (2013a). Reassessment of the unique mode of binding between angiotensin II type 1 receptor and their blockers. *PLoS ONE* 8, e79914.
- Miura, S., Okabe, A., Matsuo, Y., Karnik, S.S., and Saku, K. (2013b). Unique binding behavior of the recently approved angiotensin II receptor blocker azilsartan compared with that of candesartan. *Hypertens. Res.* 36, 134–139.
- Mozsolits, H., Unabia, S., Ahmad, A., Morton, C.J., Thomas, W.G., and Aguilar, M.I. (2002). Electrostatic and hydrophobic forces tether the proximal region of the angiotensin II receptor (AT1A) carboxyl terminus to anionic lipids. *Biochemistry* 41, 7830–7840.
- Murshudov, G.N., Vagin, A.A., and Dodson, E.J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D Biol. Crystallogr.* 53, 240–255.
- Oliveira, L., Costa-Neto, C.M., Nakaie, C.R., Schreier, S., Shimuta, S.I., and Paiva, A.C. (2007). The angiotensin II AT1 receptor structure-activity correlations in the light of rhodopsin structure. *Physiol. Rev.* 87, 565–592.
- Ramanath, V.S., Oh, J.K., Sundt, T.M., 3rd, and Eagle, K.A. (2009). Acute aortic syndromes and thoracic aortic aneurysm. *Mayo Clin. Proc.* 84, 465–481.
- Ramchandran, R., Takezako, T., Saad, Y., Stull, L., Fink, B., Yamada, H., Dikarov, S., Harrison, D.G., Moravec, C., and Karnik, S.S. (2006). Angiotensinergic stimulation of vascular endothelium in mice causes hypotension, bradycardia, and attenuated angiotensin response. *Proc. Natl. Acad. Sci. USA* 103, 19087–19092.
- Sano, T., Ohyama, K., Yamano, Y., Nakagomi, Y., Nakazawa, S., Kikyo, M., Shirai, H., Blank, J.S., Exton, J.H., and Inagami, T. (1997). A domain for G protein coupling in carboxyl-terminal tail of rat angiotensin II receptor type 1A. *J. Biol. Chem.* 272, 23631–23636.
- Shenoy, S.K., and Lefkowitz, R.J. (2005). Angiotensin II-stimulated signaling through G proteins and beta-arrestin. *Sci. STKE* 2005, cm14.
- Takezako, T., Gogonea, C., Saad, Y., Noda, K., and Karnik, S.S. (2004). “Network leaning” as a mechanism of insurmountable antagonism of the angiotensin II type 1 receptor by non-peptide antagonists. *J. Biol. Chem.* 279, 15248–15257.
- Tan, Q., Zhu, Y., Li, J., Chen, Z., Han, G.W., Kufareva, I., Li, T., Ma, L., Fenalti, G., Li, J., et al. (2013). Structure of the CCR5 chemokine receptor-HIV entry inhibitor maraviroc complex. *Science* 341, 1387–1390.
- Thomas, W.G., Baker, K.M., Motel, T.J., and Thekkumkara, T.J. (1995). Angiotensin II receptor endocytosis involves two distinct regions of the cytoplasmic tail. A role for residues on the hydrophobic face of a putative amphipathic helix. *J. Biol. Chem.* 270, 22153–22159.
- Thomas, W.G., Pipolo, L., and Qian, H. (1999). Identification of a Ca²⁺/calmodulin-binding domain within the carboxyl-terminus of the angiotensin II (AT1A) receptor. *FEBS Lett.* 455, 367–371.
- Thompson, A.A., Liu, W., Chun, E., Katritch, V., Wu, H., Vardy, E., Huang, X.P., Trapella, C., Guerrini, R., Calo, G., et al. (2012). Structure of the nociceptin/orphanin FQ receptor in complex with a peptide mimetic. *Nature* 485, 395–399.
- Totrov, M., and Abagyan, R. (1997). Flexible protein-ligand docking by global energy optimization in internal coordinates. *Proteins*, 215–220.
- Unal, H., and Karnik, S.S. (2014). Constitutive activity in the angiotensin II type 1 receptor: discovery and applications. *Adv. Pharmacol.* 70, 155–174.
- Unal, H., Jagannathan, R., Bhat, M.B., and Karnik, S.S. (2010). Ligand-specific conformation of extracellular loop-2 in the angiotensin II type 1 receptor. *J. Biol. Chem.* 285, 16341–16350.
- Unal, H., Jagannathan, R., and Karnik, S.S. (2012). Mechanism of GPCR-directed autoantibodies in diseases. *Adv. Exp. Med. Biol.* 749, 187–199.
- Weierstall, U., James, D., Wang, C., White, T.A., Wang, D., Liu, W., Spence, J.C., Bruce Doak, R., Nelson, G., Fromme, P., et al. (2014). Lipidic cubic phase

- injector facilitates membrane protein serial femtosecond crystallography. *Nat. Commun.* 5, 3309.
- Whalen, E.J., Rajagopal, S., and Lefkowitz, R.J. (2011). Therapeutic potential of β -arrestin- and G protein-biased agonists. *Trends Mol. Med.* 17, 126–139.
- White, T.A., Kirian, R.A., Martin, A.V., Aquila, A., Nass, K., Barty, A., and Chapman, H.N. (2012). CrystFEL: a software suite for snapshot serial crystallography. *J. Appl. Cryst.* 45, 335–341.
- Wisler, J.W., Xiao, K., Thomsen, A.R., and Lefkowitz, R.J. (2014). Recent developments in biased agonism. *Curr. Opin. Cell Biol.* 27, 18–24.
- Wu, B., Chien, E.Y., Mol, C.D., Fenalti, G., Liu, W., Katritch, V., Abagyan, R., Brooun, A., Wells, P., Bi, F.C., et al. (2010). Structures of the CXCR4 chemokine GPCR with small-molecule and cyclic peptide antagonists. *Science* 330, 1066–1071.
- Wu, H., Wacker, D., Mileni, M., Katritch, V., Han, G.W., Vardy, E., Liu, W., Thompson, A.A., Huang, X.P., Carroll, F.I., et al. (2012). Structure of the human κ -opioid receptor in complex with JDTic. *Nature* 485, 327–332.
- Xia, Y., and Kellems, R.E. (2013). Angiotensin receptor agonistic autoantibodies and hypertension: preeclampsia and beyond. *Circ. Res.* 113, 78–87.
- Zaman, M.A., Oparil, S., and Calhoun, D.A. (2002). Drugs targeting the renin-angiotensin-aldosterone system. *Nat. Rev. Drug Discov.* 1, 621–636.
- Zhou, C.C., Zhang, Y., Irani, R.A., Zhang, H., Mi, T., Popek, E.J., Hicks, M.J., Ramin, S.M., Kellems, R.E., and Xia, Y. (2008). Angiotensin receptor agonistic autoantibodies induce pre-eclampsia in pregnant mice. *Nat. Med.* 14, 855–862.

Structural Snapshots of Actively Translating Human Ribosomes

Graphical Abstract



Authors

Elmar Behrmann, Justus Loerke, ..., Patrick Scheerer, Christian M.T. Spahn

Correspondence

christian.spahn@charite.de

In Brief

Multiparticle cryo-EM analysis reveals 11 distinct functional states from a native actively translating human polysomal sample, providing insights into the configuration of the human ribosome at near-atomic resolution and highlighting the functional importance of both rigid and flexible regions.

Highlights

- Structural analysis of actively translating ribosomes
- Identification of significantly populated states at close to *in vivo* conditions
- Functional states feature localized chemical and conformational heterogeneity
- Human 80S ribosome map at near-atomic resolution reveals native interactions

Accession Numbers

5AJ0



Structural Snapshots of Actively Translating Human Ribosomes

Elmar Behrmann,^{1,6,7} Justus Loerke,^{1,6} Tatyana V. Budkevich,¹ Kaori Yamamoto,¹ Andrea Schmidt,^{1,2} Pawel A. Penczek,³ Matthijn R. Vos,⁴ Jörg Bürger,¹ Thorsten Mielke,^{1,5} Patrick Scheerer,^{1,2} and Christian M.T. Spahn^{1,*}

¹Institut für Medizinische Physik und Biophysik, Charité-Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany

²Institut für Medizinische Physik und Biophysik, AG Protein X-Ray Crystallography, Charité-Universitätsmedizin Berlin, Charitéplatz 1, 10117 Berlin, Germany

³Department of Biochemistry and Molecular Biology, The University of Texas Medical School, 6431 Fannin MSB 6.220, Houston, TX 77054, USA

⁴FEI Company, Nanoport Europe, Achtseweg Noord 5, 5651 GG Eindhoven, the Netherlands

⁵Max-Planck Institut für Molekulare Genetik, Ihnestr. 63-73, 14195 Berlin, Germany

⁶Co-first author

⁷Present address: Research Group Structural Dynamics of Proteins, Center of Advanced European Studies and Research (caesar), Ludwig-Erhard-Allee 2, 53175 Bonn, Germany

*Correspondence: christian.spahn@charite.de

<http://dx.doi.org/10.1016/j.cell.2015.03.052>

SUMMARY

Macromolecular machines, such as the ribosome, undergo large-scale conformational changes during their functional cycles. Although their mode of action is often compared to that of mechanical machines, a crucial difference is that, at the molecular dimension, thermodynamic effects dominate functional cycles, with proteins fluctuating stochastically between functional states defined by energetic minima on an energy landscape. Here, we have used cryo-electron microscopy to image ex-vivo-derived human polysomes as a source of actively translating ribosomes. Multiparticle refinement and 3D variability analysis allowed us to visualize a variety of native translation intermediates. Significantly populated states include not only elongation cycle intermediates in pre- and post-translocational states, but also eEF1A-containing decoding and termination/recycling complexes. Focusing on the post-translocational state, we extended this assessment to the single-residue level, uncovering striking details of ribosome-ligand interactions and identifying both static and functionally important dynamic elements.

INTRODUCTION

At the heart of many biological processes are complex and dynamic macromolecular machines. Different from macroscopic machines, these operate intermittently rather than continuously. Because inertia is irrelevant at the nanometer scale, conformational changes are dominated by thermal forces (Frauenfelder et al., 1991; Purcell, 1977). Consequently, macromolecular machines randomly sample all conformational states available to them at a given temperature instead of passing smoothly from one functional state to the other (Frauenfelder et al., 1991). Func-

tional states represent local minima in their energy landscape, defined by energetically costly conformational changes required to transit to neighboring minima.

The ribosome is an archetypical molecular machine, synthesizing proteins based on the primary sequence information encoded in mRNA templates (Frank and Spahn, 2006; Voorhees and Ramakrishnan, 2013). The ribosome consists of a large subunit (LSU; 60S in eukaryotes) containing the peptidyl transferase center (PTC) and a small subunit (SSU; 40S in eukaryotes) containing the mRNA decoding center (DC). Together, both subunits define three distinct tRNA-binding sites in their intersubunit space, referred to as the aminoacyl (A)-site responsible for binding and decoding incoming aminoacylated tRNAs, the peptidyl (P)-site responsible for orienting the polypeptide-bearing P-site tRNA for efficient transamidation, and the exit (E)-site responsible for subsequent release of deacylated tRNA.

Protein synthesis can be divided into the four phases: initiation, elongation, termination, and recycling (Melnikov et al., 2012). Each phase comprises numerous distinct functional states and multiple large-scale intra- and inter-subunit rearrangements of the ribosome, and its ligands drive the functional cycle (Dunkle and Cate, 2010; Korostelev et al., 2008). Dynamic single-molecule distance measurements show that these rearrangements are governed by a rugged energy landscape that is shaped by translation factors (Munro et al., 2009; Petrov et al., 2011). Many functional intermediates of translation have been structurally analyzed employing both X-ray crystallography and cryo-electron microscopy (cryo-EM) (Moore, 2012; Voorhees and Ramakrishnan, 2013). The focus of these studies has been on bacterial complexes, while considerably less is known about the structures of functional states sampled by ribosomes from higher eukaryotes. Traditionally, such structural studies rely on in vitro assembled complexes and on the use of antibiotics, tRNA mimics, non-hydrolyzable nucleotide analogs, or genetic modifications in order to stall ribosomes in defined states. It is still largely unknown if or how in vitro assemblies differ from their in vivo counterparts that are assembled in the complex context of the living cell. Only by investigating samples in a

native(-like) setting can these issues be addressed. While cryo-electron tomography allows the visualization of individual, active molecular machines inside cells (Brandt et al., 2010; Myasnikov et al., 2014), its resolution is limited.

Recognizing that *in vitro* systems are not able to account for the full complexity of *in vivo* environments, we considered studying native translation intermediates by imaging ex-vivo-derived non-stalled and unmodified polysomes from human cell extracts using multiparticle cryo-EM. Polysomes are formed by actively translating ribosomes and are therefore expected to constitute a mixture of ribosomes in elongation states (Rich et al., 1963). Thus, polysomes bear the potential to study the process of translation using one single specimen and to obtain not only multiple structural snapshots of functional states from the same sample, but also to determine the native distribution of states to approximate the positions of minima on the energy landscape, if conformational and compositional heterogeneity can be overcome by particle image sorting procedures (Spahn and Penczek, 2009).

To further structural insights into the process of protein synthesis inside the living cell, we report here the structural analysis of ex-vivo-derived human polysomes using multiparticle cryo-EM. We show that a variety of functional states are significantly populated, providing critical structural insights into minima of the energy landscape of the ribosomal elongation cycle and the rate-limiting steps close to the *in vivo* situation. We also demonstrate that subunit rolling (Budkevich et al., 2014) indeed constitutes a degree of freedom sampled *in vivo*. Focusing on a larger subset of particle images, we solve the structure of the human 80S ribosome in the post-translocational state at near-atomic resolution despite conformational and compositional heterogeneity. The high-resolution cryo-EM map shows details of native interactions of the ribosome with its ligands, revealing a striking difference in the binding mode between P- and E-site tRNA binding in the unrotated state and allows identifying both static and functionally important dynamic elements.

RESULTS

Distinct Functional States Can Be Reconstructed from Human Ex-Vivo-Derived Polysomes

To preserve the *in vivo* functional states of polysomes during purification, we switched from classical sucrose-gradient centrifugation to a considerably faster gel filtration-based enrichment strategy to isolate polysomes from the cytosol of human cells (Stephens and Nicchitta, 2007). Samples were vitrified in liquid ethane with minimal delay after cell lysis and enrichment, while they were still exhibiting hallmark features expected of a polysomal sample (Brandt et al., 2010; Rich et al., 1963), such as the distinct peak pattern in a sucrose gradient (Figure 1A) and clusters of ribosomes in the raw micrographs (Figure 1B). In order to sort particle images *in silico*, we employed unsupervised multiparticle analysis (Loerke et al., 2010) that was combined with 3D variability analysis to identify regions of high conformational and/or compositional heterogeneity (Extended Experimental Procedures).

A first tier of unsupervised multiparticle refinement (Figure S1) revealed tRNA-carrying ribosomes in either classical unrotated (66% of ribosomal particle images) or rotated (34% of ribosomal

particle images) intersubunit arrangement. However, both rotated and unrotated complexes still featured localized 3D variability, indicating heterogeneity in the form of substoichiometric ligand binding. We therefore employed a second tier of unsupervised classification focusing on the heterogeneous areas to further split the data into defined functional states (Figures 1C, 2, and S2). The presence of density corresponding to the nascent chain (NC) in all complexes demonstrates that our ex-vivo-derived polysomes are functional and contain predominantly active ribosomes. This is different from a recent microsomal sample, where only ~13% of the ribosomes were found in an active state (Voorhees et al., 2014). Our approach indeed allows the structural analysis of functional ribosomal complexes derived from the native environment of the cell that all were assembled and isolated under identical conditions. The resulting maps can be regarded as snapshots of the ribosome “in midflight” (Moore, 2012) allowing key insights into *in vivo* protein synthesis.

For the rotated configuration, continued sorting revealed three distinct subpopulations. The first of these contains an A/A- and a P/E-tRNA and thus represents a rotated-1 state (Figure 2A). This structure is almost identical to the *in vitro* rotated-1 PRE state (Budkevich et al., 2011), with the A-tRNA contacting H89 and H69 and the CCA end being held in the A-site, but additionally shows a contact with the ASF (Figure 3A). The second rotated state contains A/P- and P/E-configured tRNAs (Figure 2B) similar to the *in vitro* rotated-2 PRE state (Budkevich et al., 2011) and the active fraction of microsomal porcine ribosomes (Voorhees et al., 2014). Intriguingly, the dominating rotated PRE *in vivo* corresponds to the rotated-2 PRE state with two hybrid tRNAs (Figure 1C, inset; Table S1), unlike the previous bacterial structures of the rotated 70S ribosome where only the P/E-tRNA is seen in a clear hybrid position (Agirrezabala et al., 2008; Julián et al., 2008). Unexpectedly, we observe a third rotated PRE conformation with three tRNAs in classical configurations (Figure S2A). Contacts of the A-tRNA with the LSU are similar to those of the rotated-1 state. Compared to POST, the SSU is rotated by ~8°. We conclude that this rare sub-population may represent a short-lived intermediate PRE state (PRE*), however, high flexibility of the tRNA and low resolution preclude a more detailed interpretation.

For the unrotated configuration, a second tier of sorting resulted in five subpopulations. Comparison with structures from defined *in vitro* settings (Budkevich et al., 2011, 2014) identified these subpopulations as a classical-1 PRE state, two states with an A/T-tRNA, a pre-recycling state and a POST state.

Further sorting of the classical PRE state, containing three classical A/A-, P/P-, and E/E-tRNAs (Figures 2C), in a third tier of classification shows that it consists of two complexes with different amounts of rolling (Figures S2B and S2C). For the first state, the 40S subunit is rolled by ~6° with respect to the unrotated POST configuration (Figure S2G), and the overall 80S configuration matches well that of a classical-1 configuration observed *in vitro* (Budkevich et al., 2014). A second state shows intermediate rolling of ~1°–2° with respect to the POST (Figure S2B) and may correspond to an accommodation intermediate (classical-i PRE), where the interaction of the A-site tRNA with the 80S ribosome is reminiscent of the classical-2 configuration.

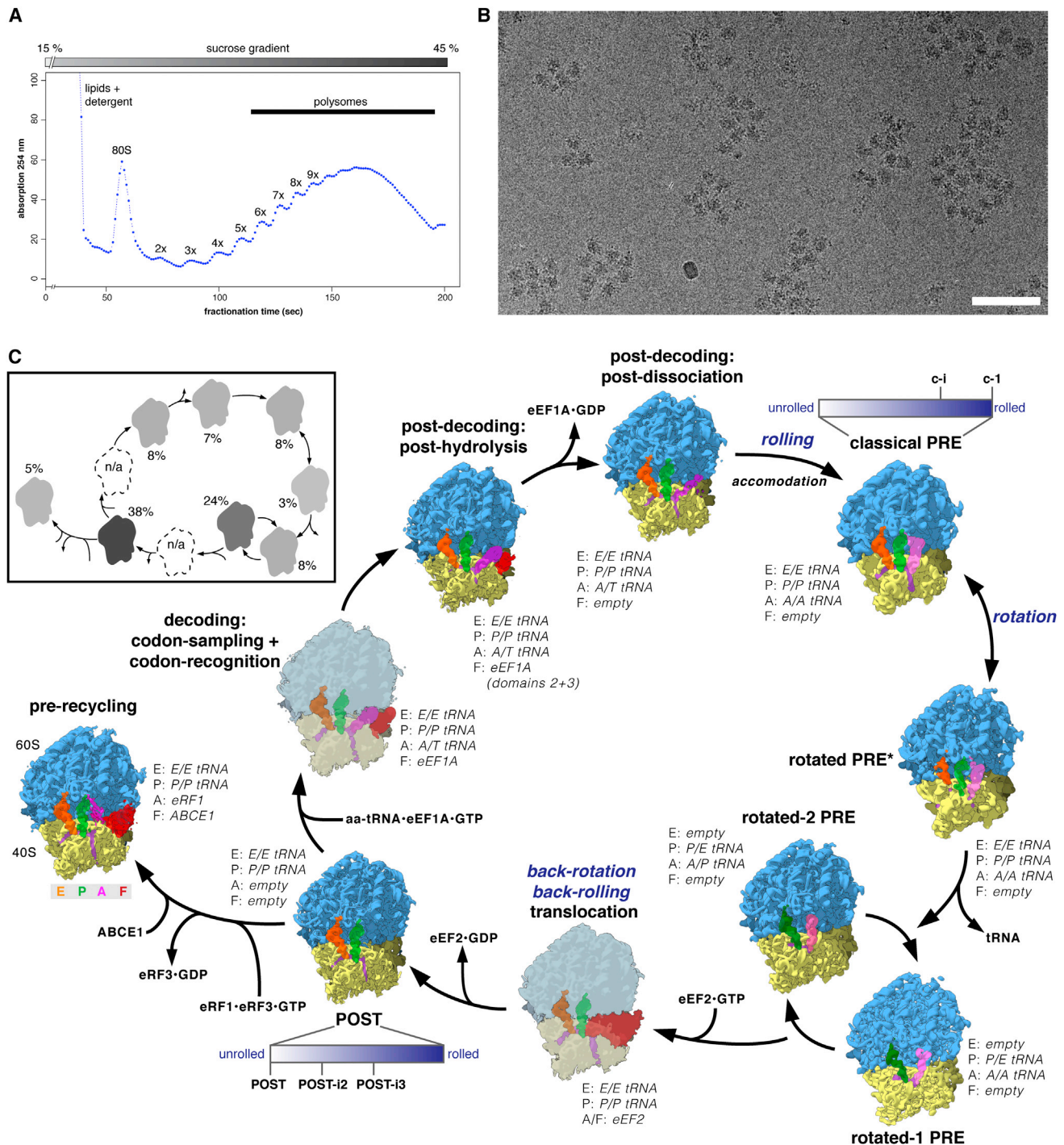


Figure 1. The Experimentally Observed Elongation Circle

(A and B) (A) Sucrose density-gradient analysis of the polysomal sample and (B) raw micrograph after size-exclusion gel-filtration. Scale bar represents 100 nm. (C) Overview of the cryo-EM maps in the framework of the elongation circle. Translocation and decoding-sampling/-recognition complexes (grayed out) were not observed experimentally and are represented by simulated maps based on published factor structures (PDB 4CXH and 2P8W). For POST and classical PRE structures with different amounts of rolling were observed, represented by the blue scale bar. All maps are filtered to 7.5 Å. Inset: relative occupancies are color-coded in grayscale.

See also Figure S1.

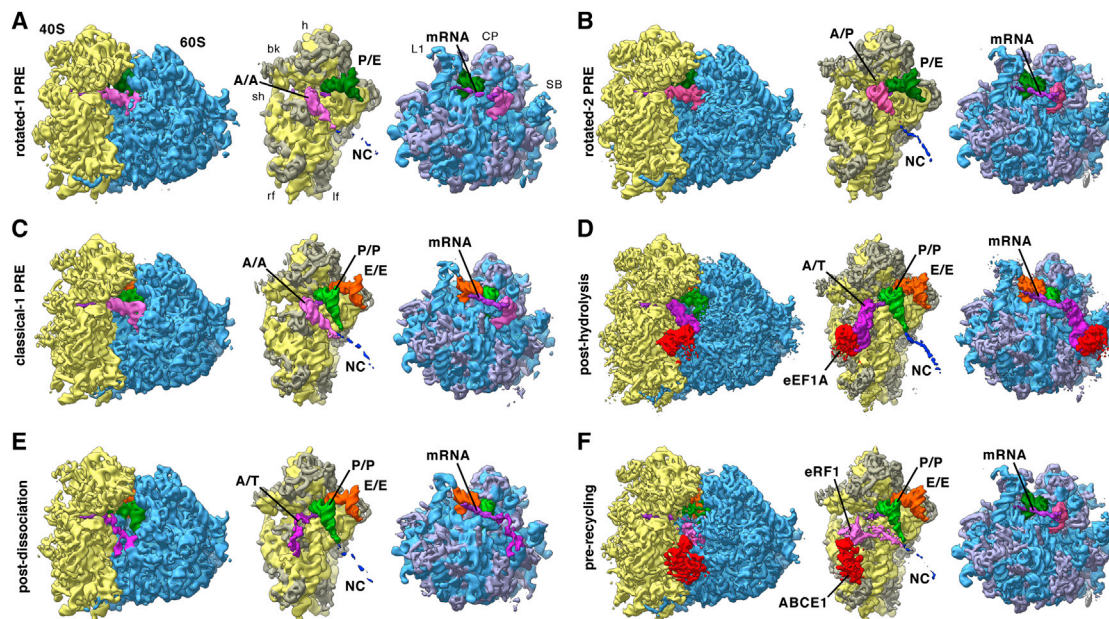


Figure 2. Functional States Reconstructed from Human Polysomes

(A–F) Cryo-EM maps filtered to their global resolution (Table S1) corresponding to (A) rotated-1 PRE, (B) rotated-2 PRE, (C) classical-1 PRE, (D) post-hydrolysis, (E) post-dissociation, and (F) pre-recycling states. For the POST state at high resolution see Figure 4. For the rotated PRE* state and states featuring intermediate amounts of rolling see Figure S2. Left: ribosomal complexes with SSU depicted in yellow and LSU in blue. Right: segmented cryo-EM maps, rotated by 80°: A/A-site tRNA (pink), A/T-site tRNA (dark pink), A/P-site tRNA (medium pink), eRF1 (pink), P-site tRNA (green), P/E-site tRNA (dark green), E-site tRNA (orange), mRNA (purple), eEF1A (red), ABCE1 (red), NC (blue), 18S RNA (yellow), 40S r-proteins (gray-yellow), 28S, 5S, 5.8S RNA (blue), and 60S r-proteins (gray-blue). See also Figure S2 and Table S1.

Interestingly, both complexes that contain classical P/P- and E/E-tRNAs and an A/T configured tRNA (Figures 2D and 2E) are different from the decoding states observed *in vitro* where eEF1A was trapped in the guanosine-5'-triphosphate (GTP) state by the non-hydrolyzable GTP analog GMPPNP (Budkevich et al., 2014). It is thus likely that the present states correspond to later decoding intermediates after GTP hydrolysis. This is corroborated by the appearance of the factor density. For the first, higher populated complex, we observe clear density for both domain III and II of eEF1A in the factor-binding site, but density corresponding to the G-domain (domain I) is highly fragmented indicating flexibility (Figure 3B). The second complex lacks significant density in the factor-binding site, although there is some density present close to the surface of the SSU where domain II of eEF1A makes contact (Figure 3C). In addition to the contacts observed for the factor-bound state, we observe a contact of the acceptor stem of the A/T-tRNA with uL14, potentially acting as a steric filter (Caulfield and Devkota, 2012), and a connection of the ASL region with the N-terminal region of eS30, which has previously been shown to also interact with eEF2 (Anger et al., 2013).

The fourth subpopulation of the unrotated states features density in the A-site that does not agree with an A-site tRNA and density in the factor-binding site different from any expected for factors involved in the elongation cycle. Comparison with *in vitro* termination complexes (des Georges et al., 2014; Preis et al., 2014) identifies this state as a pre-recycling state with bound eRF1 and ABCE1 (Figure 2F). eRF1 is in the extended

conformation with the GGQ motif of domain ce facing toward the PTC (Figure 3F). As for the *in vitro* complex (Preis et al., 2014), its NTD is fragmented. Similarly, the distal nucleotide-binding domain (NBD2) of ABCE1 is fragmented.

A major fraction of particle images of our *ex-vivo*-derived polysomes was assigned to the POST state. As POST state complexes appear to be stable in terms of conformation (Budkevich et al., 2014), we continued refinement of this subpopulation to improve its resolution. Although it has been demonstrated that near-atomic resolution maps for relative invariant parts of ribosome can be obtained, e.g., by focusing the refinement on the large ribosomal 60S subunit (Penczek et al., 2014), and composite near-complete atomic models of the eukaryotic ribosome can be constructed by combining the best resolved parts from different functional states (Voorhees et al., 2014), we tried to represent distinct complexes by a single cryo-EM map each. This was to ensure that we describe distinct functional states and are able to faithfully visualize structural links between remote functional sites (Agmon et al., 2005), e.g., tRNAs bridging the ribosomal subunits or the dynamic inter-subunit bridges (Gambashvili et al., 2000). Intriguingly, further sorting of the population representing POST state complexes revealed a degree of freedom with regard to the presence of subunit rolling (Figures S2D–S2F). While the majority of particles did not show any subunit rolling and was refined to high resolution, we obtained two additional populations (POST-i2 and POST-i3) with ~1° and ~3° of subunit rolling, respectively (Figures S2D and S2E). Due to the limited resolution of these two states, we cannot discern

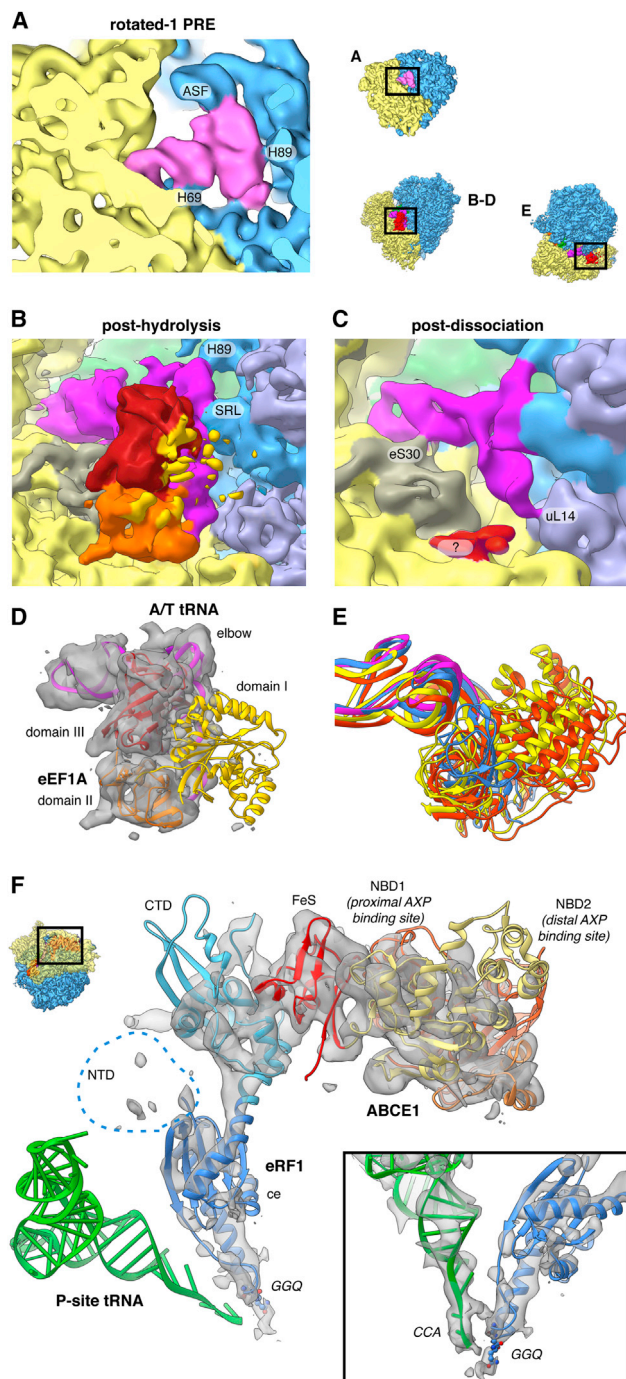


Figure 3. Imaging Ex-Vivo-Derived Polysomes Allows the Visualization of Transient States

(A–E) Close-up view of the rotated-1 PRE state A-site tRNA (A) and the post-decoding states (B–E). (B and D) Only domain III (red) and domain II (orange) of eEF1A feature strong density, while domain I (yellow) is fragmented. The A/T-tRNA (dark pink) elbow is connected to the SRL and H89. (C) For the post-dissociation state additional contacts with eS30 and uL14 are visible. A fragmented density of unclear origin is shown in red. (E) 18S RNA-based overlay of decoding-sampling (yellow), decoding-recognition (orange), post-decoding (blue) and post-dissociation (pink) models for eEF1A and the A/T-tRNA elbow.

whether they constitute true energetic minima of subunit rolling, or encompass a continuous band of subunit rolling.

In total, our three-tier multiparticle refinement strategy enabled us to identify and visualize 11 distinct functional states of translating human ribosomes, the majority corresponding to elongation states. All are resolved to sub-nanometer resolution or better (Figures S3A–S3D; Table S1) and show robust ligand-densities (Figures 2 and S2).

Structure of the Native Human POST Complex at Near-Atomic Resolution

After refining the largest POST population of 313,321 particle images (16% of the total data set) separately, we obtained a highly improved cryo-EM map for the POST state with a global resolution of 4.0 Å using the 0.5 Fourier shell correlation (FSC) criterion, whereas the 0.143 FSC criterion suggests that the map is equivalent to an X-ray density map at 3.5 Å resolution (Figure S3E). We corroborated this resolution estimate by a local resolution measurement that is independent of the FSC (Figure S3F). Visual inspection of the map agreed with the near-atomic resolution estimate, with the cryo-EM map (Figure 4) allowing direct observation of single-residue details for large parts of the map (Figures 4C–4F and S4A–S4D). However, intrinsically flexible expansion segments remain less defined (Figures S3G and S3H), indicating that these structural elements are uncoupled from the functional state of the ribosome. Moreover, all elements endogenously present as mixtures remain less defined, with exception of the remarkably well-resolved P-site tRNA.

The quality of the cryo-EM map in the well-ordered regions appears comparable to that of recent crystal structures of eukaryotic ribosomes (Ben-Shem et al., 2011; Klinge et al., 2011; Rabi et al., 2011), allowing us to resolve individual nucleotides with distinct densities for phosphates, bases, and sugars, as well as protein backbones with clearly visible bulky side-chains. Starting from our previous homology model (Figures S4E and S4F) of the human ribosome (Budkevich et al., 2014), we created an atomic model for the human ribosome (Tables S2 and S3) by iterating multiple rounds of (semi-)manual real-space fitting, energy minimization and geometric idealization (Extended Experimental Procedures; Table S4). The quality of the cryo-EM map allows rationalization of single point mutations compared to yeast (Figure 4E) and determination of correct residues for ambiguous protein sequences (Figure S4A). The high signal-to-noise-ratio of the ordered regions allows the visualization of individual charged ions (Figures 4F, S4B, and S4C). We tentatively assigned these as either chelated or diffuse magnesium based on comparison to known magnesium binding sites (Jenner et al., 2010) and binding motifs (Klein et al., 2004). In total, the atomic model provides a detailed inventory of protein-protein, RNA-RNA and protein-RNA interactions that define the human ribosome in the native, unrotated POST state, while previous high-resolution structures of the 80S ribosome where all solved in rotated or partially rotated conformations (Ben-Shem et al., 2011; Voorhees et al., 2014).

(F) Close-up view of the pre-recycling state showing eRF1 (shades of blue) and ABCE1 (yellow to red). Atomic models are based on Preis et al. (2014). See also Movie S1.

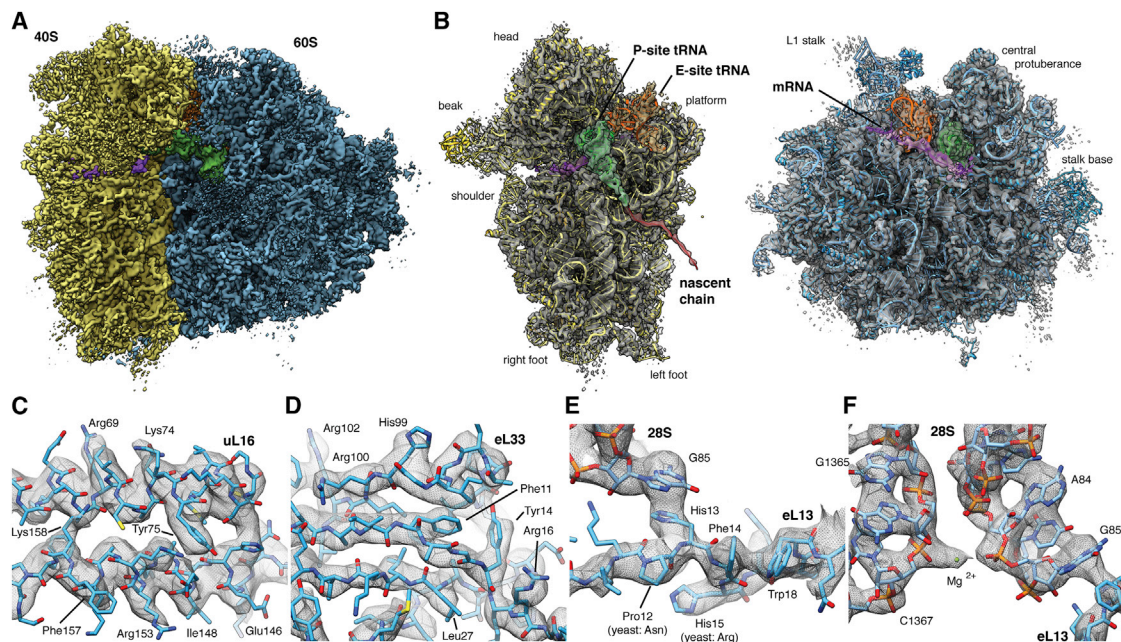


Figure 4. High-Resolution Structure of the Human Ribosome in the POST State

(A) Surface representation of the POST state cryo-EM map filtered to 3.5 Å (blue, LSU; yellow, SSU; green, P-site tRNA; orange, E-site tRNA; purple, mRNA). (B) Individual subunit maps with the corresponding atomic models. Segmented density corresponding to the NC (red) is shown filtered to 7.0 Å for clarity. Segmented maps are shown turned by 80°.

(C–F) Enlarged regions of the cryo-EM map showing well-resolved (C) alpha-helices or (D) beta-strands with individual side-chains, (E and F) strong π -stacking interactions, and (F) individual nucleotides with nearby ions.

See also [Figures S3](#) and [S4](#) and [Tables S2](#), [S3](#), and [S4](#).

Molecular Description of Eukaryotic-Specific Bridges in the Unrotated Configuration

Our present map facilitates the assessment of interactions between the ribosomal subunits via eukaryotic bridges in the classical, unrotated subunit arrangement. As the dynamic nature of the intersubunit bridges is prerequisite to support large-scale conformational changes of the ribosome, like intersubunit rotation or 40S subunit rolling, molecular knowledge of the bridges in all relevant configurations is crucial. Our high-resolution structure now validates our initial assignment of intersubunit bridges ([Budkevich et al., 2014](#)) and reveals molecular details for most of the intersubunit interactions in the POST state ([Table S5](#)). Especially, the lateral eukaryotic-specific bridges eB12 and eB13 are affected by intersubunit rearrangements. For example, the distal part of the C-terminal helix of eL19, forming bridge eB12, is displaced by up to 25 Å ([Figure 5A](#)) in comparison to the yeast crystal structures ([Ben-Shem et al., 2011](#)). Remarkably, the interaction interface with the large groove of expansion segment es6E on the 40S side is hardly affected by this: e.g., the interaction between Arg163 of eL19 and U871 (yeast U813) of es6E is maintained irrespective of the intersubunit arrangement ([Figure 5B](#)). On the opposing side of the SSU, bridge eB13 acts akin to a tethered anchor, with a flexible linker of eL24 (residues 68–85) allowing for highly similar binding positions of the C-terminal kinked “anchor” regardless of the intersubunit arrangement ([Figure 5C](#)). Different from the lateral bridges, the central eukaryotic-specific bridge eB14 comprising the highly

conserved peptide eL41 is largely unaffected by intersubunit rearrangements ([Figure 5D](#)). Interestingly, eL41 folds into a linear alpha-helix reminiscent of an axle that binds into a “socket” formed between several rRNA helices of the SSU. Potentially, eB14 could thus help defining the motion center of 40S rolling and rotation.

Interactions of the Ribosome with a Classical P-Site tRNA

Although ex-vivo-derived polysomes contain a mixture of all endogenous tRNAs, the P-site tRNA density is almost completely defined to high resolution. Exceptions localize to regions with known structural variability, especially the variable loop and the D-stem loop ([Figure 6A](#)) ([Giegé et al., 2012](#)). The well-resolved density of the P-site tRNA implies that at least for the vast majority of endogenous tRNAs a single conformation is enforced by the P-site binding pocket. Comparison with crystal structures of *Thermus thermophilus* ribosomes ([Selmer et al., 2006](#)) demonstrates a striking level of structural conservation. Still, we note a direct interaction between the C-terminal Arg146 of uS9 and the tRNA at positions 33 and 35 ([Figure 6B](#)) different from bacterial structures ([Selmer et al., 2006](#)). It is apparently a swap of Lys145 for a tyrosine compared to bacteria ([Figures 6C](#) and [S5](#)) that changes the electrostatic situation at the C terminus, promoting the direct contact.

At the tRNA elbow the P-site loop around Arg64 of uL5, monitoring P-site occupancy ([Rhodin and Dinman, 2010](#)),

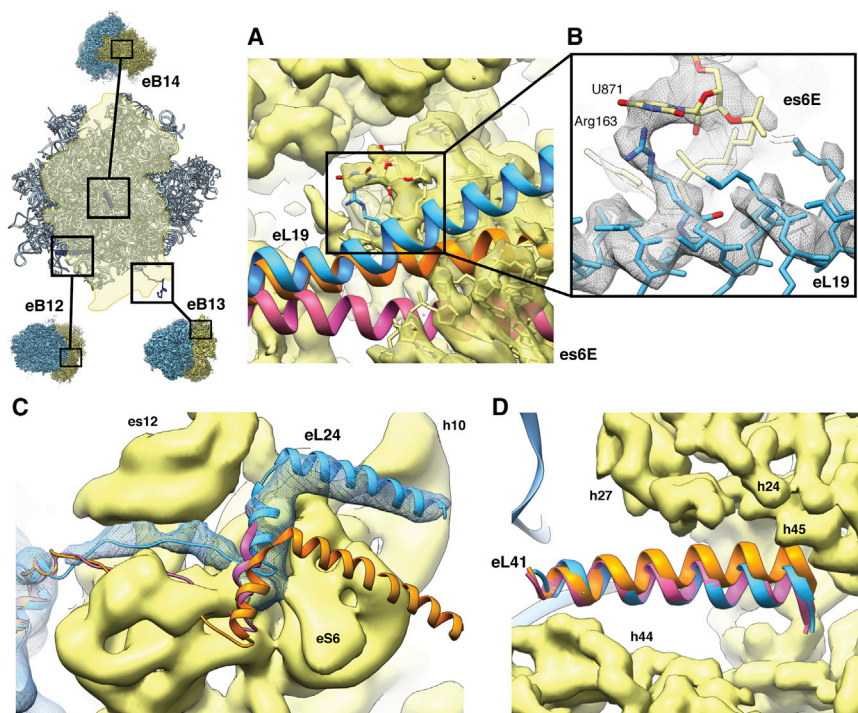


Figure 5. Eukaryotic-Specific Bridges eB12, eB13, and eB14 Are Differentially Affected by Intersubunit Rotation

Comparison of yeast LSU crystal structures (ribosome A, orange; ribosome B, purple) (Ben-Shem et al., 2011) with the present unrotated human atomic model (blue). The orientation aid illustrates the orientation of the 80S in each panel. (A and B) In the unrotated state, the C-terminal helix of eL19 forming eB12 is bent compared to the yeast structures (A), however, virtually identical interactions are observed between eL19 and es6E (B).

(C) To visualize the flexible linker tethering the C-terminal kinked “anchor” of eL24 forming eB13, density is shown filtered to 7.0 Å. Despite a strong displacement of the “anchor,” its overall shape remains highly similar.

(D) The central bridge eB14 is hardly affected by intersubunit rearrangements.

See also Table S5.

Visualization of Chemically Heterogeneous NC and mRNA

Despite the chemical heterogeneity of the NC, we observe a continuous density extending from the P-site tRNA into the exit tunnel, most likely repre-

sents the first five to six residues of the NC. For the amino acid connected to the CCA-end of the P-tRNA, a smeared-out density bulge may represent a mixture of all endogenous side chains (Figure 6E). Limiting the resolution to 7 Å allowed us to trace the path of the NC through the complete LSU (Figure S6A).

Similarly to the NC, the mRNA is expected to contain mixtures of nucleotides at each position. Although the mRNA density is largely fragmented, we are able to trace the path of approximately 28 nucleotides when limiting the resolution to 7 Å (Figure S6B). At the A-site on the SSU, only A1824 (A1492) and A1825 (A1493) of h44 are disordered (Figure S6C), most likely sampling flipped-in and -out positions as no A-site tRNA is present. At the P-site, despite the heterogeneity of codons, the mRNA density is well resolved at full resolution (Figure S6D), resembling the situation observed for the P-site tRNA. At the E-site, individual bases of the mRNA are still defined, especially at codon positions −1 and −2, but the wobble position −3 is partially fragmented (Figure S6D).

contacts the T-stem loop (Figure 6B). Moreover, the C-terminal Phe106 of eL44 establishes an additional tentative contact with the tRNA via aromatic stacking. Intriguingly, uL5 interacts through the neighboring Asn65 side chain with the backbone carbonyl oxygen of Gly101 in eL44, suggesting a concerted action of both proteins at the P-site (Figure 6B). Contrary to our in vitro ribosomal complex (Budkevich et al., 2014), we do not observe a direct interaction of uL16 with position 1 of the P-site tRNA. Rather, at high resolution clear density for the residues comprising the tip of the loop is lacking, implying heterogeneity or flexibility (Figure 6B, inset). This agrees best with a transient interaction in the POST state, suggesting release after guiding peptidyl-tRNA from the A/P hybrid position to the classical P/P position in the POST state (Budkevich et al., 2011).

At the PTC, superimposing structures of *T. thermophilus* containing three tRNAs (Selmer et al., 2006) and *Haloarcula marismortui* LSU containing tRNA-mimics (Schmeing et al., 2003) with our model emphasizes the high conservation between the three domains of life. The backbone atoms of the PTC residues superimpose with a root-mean-square deviation of 0.77 Å and 0.84 Å, respectively. We observe identical interactions of the P-tRNA acceptor stem where residues C74, C75, and A76 stack (Figure 6D). C74 and C75 furthermore form Watson-Crick base pairs with residues G4159 (*Escherichia coli* numbering G2252; *E. coli* numbering will be given in brackets in the following) and G4158 (G2251) of the 28S RNA P loop, respectively. The terminal A76 is stabilized by interaction with A4359 (A2451). The well-described A-minor interaction between A76 and the C3880 (C2063) and A4358 (A2450) pair (Selmer et al., 2006) is present in native human ribosomes.

presenting the first five to six residues of the NC. For the amino acid connected to the CCA-end of the P-tRNA, a smeared-out density bulge may represent a mixture of all endogenous side chains (Figure 6E). Limiting the resolution to 7 Å allowed us to trace the path of the NC through the complete LSU (Figure S6A).

Similarly to the NC, the mRNA is expected to contain mixtures of nucleotides at each position. Although the mRNA density is largely fragmented, we are able to trace the path of approximately 28 nucleotides when limiting the resolution to 7 Å (Figure S6B). At the A-site on the SSU, only A1824 (A1492) and A1825 (A1493) of h44 are disordered (Figure S6C), most likely sampling flipped-in and -out positions as no A-site tRNA is present. At the P-site, despite the heterogeneity of codons, the mRNA density is well resolved at full resolution (Figure S6D), resembling the situation observed for the P-site tRNA. At the E-site, individual bases of the mRNA are still defined, especially at codon positions −1 and −2, but the wobble position −3 is partially fragmented (Figure S6D).

POST State Ribosomes Undergo Stable Interactions with E-Site tRNA

We observe distinct density corresponding to a tRNA bound to the E-site, demonstrating that the E-site tRNA in the POST state is at least stable enough to survive gel filtration. Different from the well-resolved P-site tRNA, the bulk of the E-site tRNA density is fragmented (Figure 7A), suggesting that the E-site allows for a more relaxed binding. While we discern no direct interactions with the SSU, on the LSU side we observe a delocalized interaction between the tRNA elbow and the L1 stalk (Figure S7A). The full definition

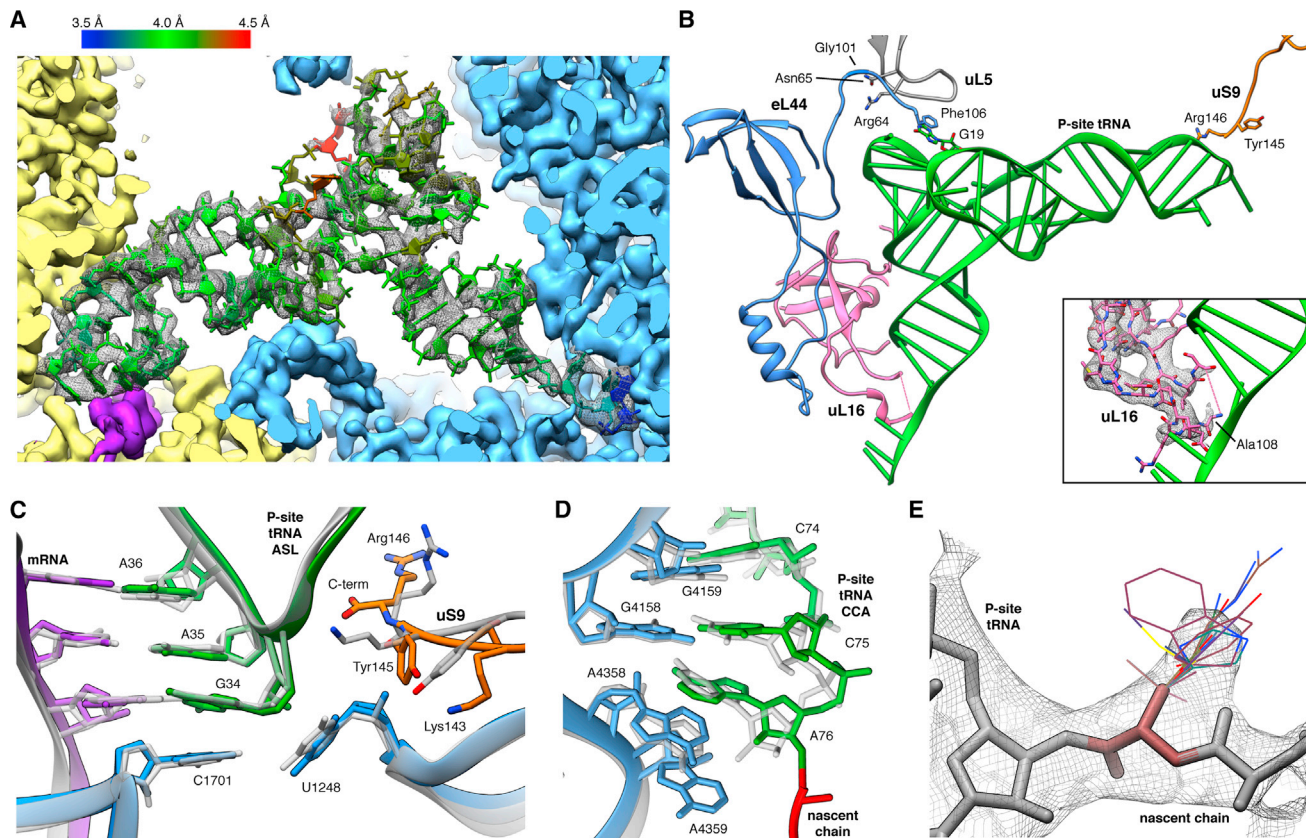


Figure 6. The P-Site tRNA Is Defined Despite Chemical Heterogeneity

(A) Cryo-EM map (mesh) and atomic model of the P-site tRNA (colored by local resolution as determined by ResMap). SSU, yellow; LSU, blue; mRNA density, purple.

(B) Key interactions of the P-site tRNA with its binding pocket on LSU and SSU. The inset shows the fragmented density of the uL16 P-site loop (mesh).

(C and D) Comparison between prokaryotic (transparent) and human (color) (C) ASL and (D) PTC. Atomic models of the prokaryotic (PDB 2J00) and the eukaryotic LSU were aligned based on the LSU rRNA.

(E) Cryo-EM map (mesh) of A76 and the first residues of the NC. Stick representations depict the most abundant rotamers of each amino acid with the exception of phenylalanine and tyrosine, where less abundant rotamers are depicted, and proline, which is not shown.

See also Figures S5 and S6.

of the CCA-end of the E-site tRNA (Figure 7B) implies that the fragmented appearance of the majority of the E-site density is not caused by substoichiometric occupancy but by conformational heterogeneity, which in turn may be caused by small differences among different tRNA species and/or flexibility/mobility. This is corroborated by the full presence of the whole E-site tRNA when the resolution of the map is limited to 7 Å (Figure S7B).

Unlike the body of the E-site tRNA, the acceptor stem features a well-defined density due to a contact with the 28S RNA at U3686 and G3711 and strong interactions of the CCA-end with the LSU (Figure 7B). As observed for prokaryotes and archaea, A76 is tightly packed in a sandwich between G4332 (G2421) and G4333 (C2422) forming a binding pocket excluding aminoacylated CCA-ends (Schmeing et al., 2003). In addition, C75 interacts by π -stacking with Tyr41 of eukaryote-specific eL44, while in prokaryotes C75 and C74 are stabilized by internal nucleotide stacking (Selmer et al., 2006). Converse to our

findings, a preceding structure of the *H. marismortui* LSU in complex with a CCA tri-nucleotide suggested that instead of π -stacking Arg40 and Gly57 of eL44 provide additional stabilization of the E-site CCA-end (Schmeing et al., 2003). However, sequence alignment highlights that the eL44 interaction observed in our structure likely corresponds to the general case in eukaryotes as *H. marismortui* harbors an eL44 sequence unique to halobacter. Other archaea and eukarya feature a conserved Tyr or Phe at position 41 (Figure S5). Superimposing the *H. marismortui* crystal structure onto our model indicates that *H. marismortui* Phe52 occupies almost the same position as Tyr41 in human and could potentially rearrange under physiological conditions to interact with C75. Superposition also reveals that the loop extension of eL44 between Phe56 and Thr62, which widens upon E-site binding, is shifted up to 4 Å when compared to *H. marismortui* (Schmeing et al., 2003), most likely due to steric clashes between tRNA and the loop extension of eL44 (Figures 7C–7E).

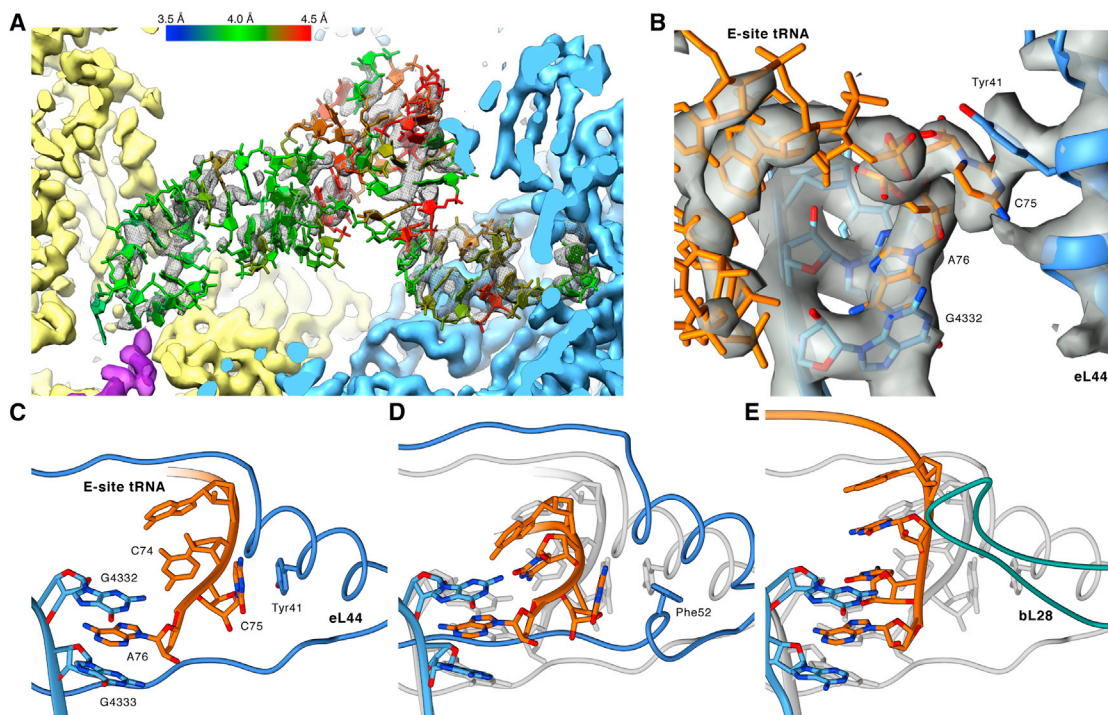


Figure 7. Native POST State Ribosomes Contain an E-Site tRNA

(A) Cryo-EM map (mesh) and atomic model of the E-site tRNA (colored by local resolution as determined by ResMap).

(B) Close-up on the cryo-EM map (transparent gray) of the CCA-end of the E-site tRNA (orange) and surrounding LSU elements (blue).

(C–E) Comparison between human, archaeal, and bacterial E-site CCA-ends. Atomic models were aligned based on the LSU rRNA and depict (C) the human, (D) the *H. marismortui*, and (E) the *T. thermophilus* CCA-end. In (D) and (E), the human model is shown in transparent gray for comparison.

See also Figures S5, S6, and S7.

DISCUSSION

The Multi-Tiered Landscape of Translation Elongation inside the Cell

Cytosolic polysomes comprise actively translating ribosomes sampling a large variety of functional states. Despite this heterogeneity, our data-driven *in silico* sorting scheme allowed us to uncover 11 distinct states and visualize them with at least sub-nanometer resolution (Figure 1). We cannot rule out that certain more transient states were lost during our purification procedure. However, the fact that the vast majority of particle images were assigned to bona fide functional states and the richness in functional states covering most of the ribosomal elongation cycle shows that the stable intermediates have been preserved. Comparing the relative SSU-LSU configurations among all states (Figures S2G–S2I), we note that in line with our preceding *in vitro* studies aimed at unveiling key transitory states of mammalian elongation (Budkevich et al., 2011, 2014) intersubunit rotation and eukaryotic-specific 40S subunit rolling are indeed conformational modes present *in vivo*.

However, we observe neither complexes with significant head swiveling nor complexes containing eEF2. Non-stalled eEF2 has been observed in a subpopulation of *in vitro* assembled complexes (Budkevich et al., 2014) and on the populations isolated from mammalian cells that showed inactive 80S complexes

(Anger et al., 2013; Voorhees et al., 2014). Thus, eEF2 binding to actively translating 80S ribosomes appears less stable implying that translocation states are short-lived intermediates *in vivo*. Kinetic (Guo and Noller, 2012) and structural studies (Spahn et al., 2004; Ratje et al., 2010) have linked head swiveling to EF-G/eEF2 containing translocation intermediates. The simultaneous absence of eEF2 and significant head swiveling in our present elongation intermediates supports such a model.

While not sufficient to map the exact topology of the energy landscape of elongation, our structural description allows assessment of the distribution of the more stable functional states of active polysomes in a native setting (Figure 1C, inset). This distribution correlates with the relative energetic stability of each state in comparison to the most stable one (Fischer et al., 2010; Frank, 2013). We note that neither is the unrolled configuration exclusive to the POST state nor the rolled configuration to the classical PRE states. Rather, rolling appears to be a spontaneous movement where the presence of an A-site tRNA shifts the equilibrium toward the rolled classical PRE-state.

From the local heterogeneity uncovered by our 3D variability analysis, it is moreover evident that the energetic minimum corresponding to each state can potentially be split into finer substates, given proper sorting of the particle images. While functionally defined, each of these states features localized heterogeneity due to structural elements that are more flexible—in line with the assumption of a multi-tiered hierarchical energy

landscape governing elongation (Munro et al., 2009) and protein activity in general (Frauenfelder et al., 1991).

While after 3D variability analysis and focused classification a major part of the POST complex can be resolved at single-residue resolution, there are still regions exhibiting a fragmented appearance. This localized disorder can be regarded as structural evidence for the finely split sub-valleys of the energy landscape. The most striking example is the highly localized disorder of the bases A1824 (A1492) and A1825 (A1493) involved in A-site decoding (Demeshkina et al., 2012; Ogle et al., 2003) in the otherwise well-ordered h44 (Figure S6C). Both bases are likely sampling both flipped-in and -out positions as no A-site tRNA is present. Such flexibility has been predicted from molecular dynamics simulations (Sanbonmatsu, 2006), but is in contrast to recent X-ray crystallization data from bacterial complexes with P-site tRNA and mRNA showing a partially preformed DC with A1493 stably flipped-out and A1492 stably flipped-in by stacking with A1913 (Demeshkina et al., 2012). Still, our findings can easily be reconciled with the known kinetics of decoding by assuming that the empty DC, behaving akin to a liquid unstructured region, rigidifies upon codon-anticodon interaction. In this “flow-fit” model, the mobile decoding bases allow fast sampling of the codon-anticodon duplex with the cognate tRNA inducing a stronger binding interface and thus rigidifying the DC with higher probability and higher speed. This matches well the experimental observation that both cognate and near-cognate tRNAs bind to the ribosome with the identical rates, but that cognate tRNAs dissociate not only more slowly from the DC compared to near-cognate tRNAs, but also exhibit faster rates for the forward reactions, i.e., GTPase activation (Pape et al., 1999; Geggier et al., 2010). Thus, the observed disorder of the bases A1824 (A1492) and A1825 (A1493) in an otherwise highly ordered environment exemplifies the potential biological importance of mobile, more “liquid” regions (Dyson, 2011) and demonstrates the potential of visualizing macromolecular machine at near-atomic resolution in a solution-like state under near-physiological conditions.

Native Proofreading Complexes Highlight Rate-Limiting Steps

The presence of ribosomal decoding complexes in polysomes is not immediately expected, as decoding complexes are believed to be short-lived intermediates. Accordingly, structural investigations of ribosome-bound ternary complexes rely on the use of non-hydrolyzable GTP analogs or antibiotics to inhibit the transition of EF-Tu/eEF1A from the GTP to the GDP conformation (Budkevich et al., 2014; Schmeing et al., 2009; Schuette et al., 2009; Villa et al., 2009; Voorhees et al., 2010). The visualization of a non-stalled ternary complex on the human ribosome, and of a second ribosomal decoding complex containing only A/T-tRNA but no factor, demonstrates the power of our approach that aims to derive structures of functional complexes from the on-going functional cycle instead of isolated functional complexes. Under steady-state/multi turnover conditions even shorter-lived states may be significantly populated, as they are constantly replenished.

Comparing the ternary complex observed from polysomes to decoding-sampling and decoding-recognition complexes from

our preceding in vitro study with GMPPNP-stalled eEF1A (Budkevich et al., 2014), we note significant differences. First, the elbow of the A/T-tRNA has already released the stalk base and appears more strongly bound to the sarcin-ricin loop (SRL, H95) instead (Figure 3B). Second, there is relevant disorder of domain I containing the GTP-binding pocket (Figure 3D). These differences can be readily reconciled by the notion that both sets of complexes represent different states along the pathway of tRNA selection. For the in vitro complexes eEF1A was trapped in the GTP state by GMPPNP and accordingly the complexes were observed in the initial phase of decoding before GTP hydrolysis (Budkevich et al., 2014). As in our present study, the chemical step is not inhibited, and based on the trajectory of structural changes from decoding-sampling to decoding-recognition (Budkevich et al., 2014) to the present complex (Figure 3E; Movie S1), we infer that it corresponds to an aminoacyl(aa)-tRNA•eEF1A•GDP ternary complex in the post-hydrolysis/proofreading state before the eEF1A dissociation and accommodation steps. This implies that the transition to the GDP-induced conformation of the factor and release of aa-tRNA from eEF1A•GDP do not occur immediately upon SRL-promoted GTP hydrolysis, but with a significant delay. This is in excellent agreement with kinetic studies in the bacterial system where GTP hydrolysis has been shown to be a very fast step, whereas tRNA accommodation and especially EF-Tu dissociation are rate-limiting during tRNA selection (Pape et al., 1998). Furthermore, our results rationalize recent studies on the formation and turnover of bacterial EF-Tu•GXP•EF-Ts•aa-tRNA quarternary complexes (Burnett et al., 2013, 2014), which proposed a novel role of EF-Ts in promoting release of aa-tRNA from EF-Tu•GDP. The presence of a second proofreading state containing A/T-tRNA, but lacking density for eEF1A, implies that also tRNA accommodation constitutes a second slow step after eEF1A dissociation, in agreement with the hypothesis that necessary conformational changes in the tRNA elbow to allow A/T to A/A transition resemble a stochastic trial-and-error process and not a concerted pathway (Whitford et al., 2010; Geggier et al., 2010). We believe that from both post-hydrolysis states near-cognate tRNA can be rejected in line with the concept of kinetic proofreading (Hopfield, 1974).

A Stably Occupied E-Site Is an In Vivo Feature of Elongating Human Ribosomes

The properties of the E-site tRNA in the bacterial system have been controversially discussed for decades and it is still not generally agreed on whether the E-site is only transiently occupied directly after translocation, or whether the E-site tRNA is released at latter stages, i.e., during A-site occupation (Wilson and Nierhaus, 2006). Our maps now show an occupied E-site also in functional states subsequent to the POST state, with three tRNAs present in the classical PRE, post-decoding and PRE* complexes. Solely for the rotated-1 and rotated-2 PRE complexes do we observe only two tRNAs. Remarkably, during human translation elongation the 60S E-site appears to be always occupied by the CCA-end of either an E/E- or a P/E-tRNA. While it can be argued that the presence of an E-site tRNA on in vitro-assembled complexes is due to the excess of deacylated tRNA used for technical reasons (Budkevich et al.,

2011, 2014), this argument falls short for the present ex-vivo-derived complexes. Thus, our findings strongly suggest that stable E-site occupation in all but the rotated-1 and rotated-2 PRE states is an in vivo feature of the human system.

Native PRE Complexes Pinpoint the Release of E-Site tRNA

Given that our structures allow us to trace the transition from an unrotated PRE state with three classical tRNAs to the rotated PRE states with two tRNAs in either classical or hybrid configuration, and especially due to the observation of a rotated PRE* subpopulation with A/A-, P/P-, and E/E-tRNAs, we can pinpoint the release of E-site tRNA during human translation elongation. Apparently, it is the rotation between the two subunits that critically destabilizes the E-site, leading to subsequent release of the E-site tRNA (Figure 1C). Still, while our structures definitely support the existence of the aforementioned pathway, it has to be noted that this observation does not preclude the possibility of parallel alternative pathways where either E-site tRNA is released concomitantly with intersubunit rotation, or even before rotation.

Conclusions

Our study and others have recently demonstrated that cryo-EM has opened up the way to study the structure of the ribosome at high resolution unconstrained by a crystal lattice. Furthermore, by relying on thorough in silico classification, we have demonstrated that defined structures corresponding to known and hitherto unknown intermediate states of translation can be obtained from ex-vivo-derived elongating polysomes. Different from traditional “arrest and isolate” strategies, our approach has shed light on the preferred states of the human ribosome. This uncovered that while eEF2-mediated head rotation is paramount for translocation, corresponding functional states are only sparsely populated precluding visualization, corroborating the assumed short-lived nature of the translocation state. Analyzing the full spectrum of significantly populated states of elongation has not only addressed the specific point of E-site tRNA release, but also uncovered that proofreading states of tRNA selection, after codon recognition and GTP hydrolysis, can be significantly populated, implying that tRNA accommodation is indeed a slower step. Importantly, despite the high degree of complexity of a native-like polysomal sample, we were able to overcome heterogeneity using a data-driven sorting scheme. This allowed us to resolve the native POST state to near-atomic resolution and thus highlight the divergent properties of P- and E-site and uncover dynamic elements in the ribosome, such as the decoding bases.

EXPERIMENTAL PROCEDURES

Additional details can be found online in the [Extended Experimental Procedures](#).

Polysome Isolation and Grid Preparation

Polysomes were prepared from the cytosolic fraction of digitonin permeabilized HEK293T cells (Hirashima and Kaji, 1970; Stephens and Nicchitta, 2007). The cytosolic fraction was further separated on a Sepharose 4B size-exclusion column, isolating polysomes as the first peak absorbing at 254 nm. Samples were immediately prepared for cryo-EM by vitrification.

Data Collection

Electron micrographs were collected automatically on an FEI Krios microscope equipped with a back-thinned Falcon II detector and on an FEI Tecnai G2 Polara equipped with a TC-F416 CMOS camera. The total data set comprised 51,282 micrographs yielding 1,823,338 particles (801,789 Krios, 1,121,549 Polara).

Data Processing

The data set was processed using incremental K-means-like procedures (Loerke et al., 2010) in SPIDER (Frank et al., 1996). Initially, the two data sets split into subsets belonging to either rotated or unrotated ribosomal complexes or to artifactual particles. Particle images belonging to the rotated PRE and unrotated POST states were separated and artifactual particle images were removed. Particle images were refined and classified further using 3D variability analysis to guide sorting. The final map of the POST state, based on 313,321 particle images (130,953 Krios, 182,368 Polara), reached a resolution of 4.0/3.5 Å. Cryo-EM density maps have been deposited with the EMDB (accession number EMD-2875, EMD-2902, EMD-2903, EMD-2904, EMD-2905, EMD-2906, EMD-2907, EMD-2908, EMD-2909, EMD-2910 and EMD-2911) and coordinates for the POST state have been deposited with the Protein Data Bank (entry code 5AJ0).

Model Building and Refinement

Initial atomic models of *H. sapiens* 40S and 60S subunits derive from our preceding study (Budkevich et al., 2014). Ligands were rebuilt based on crystal structures of prokaryotic or archaeal tRNAs. The NC poly-alanine model was built de novo. Overlapping stretches of the model were manually adjusted into the cryo-EM map as rigid bodies, followed by real space refinement and geometric idealization for well-resolved densities. Structure models were further refined and validated using crystallography tools.

ACCESSION NUMBERS

The following accession numbers for the cryo-EM density maps reported in this paper are available in the EMDB: EMD-2875, EMD-2902, EMD-2903, EMD-2904d, EMD-2905, EMD-2906, EMD-2907, EMD-2908, EMD-2909, EMD-2910, and EMD-2911. The accession number for the coordinates for the POST state reported in this paper is PDB 5AJ0.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, five tables, and one movie and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.03.052>.

AUTHOR CONTRIBUTIONS

T.V.B. and K.Y. prepared polysome samples. T.M. and M.R.V. supervised and coordinated cryo-EM data collection. J.B. performed data collection. J.L., E.B., and C.M.T.S. processed, refined, and analyzed cryo-EM data (based on refinement strategies designed by J.L., P.A.P., and C.M.T.S.). E.B., A.S., and P.S. refined atomic models. E.B., J.L., and C.M.T.S. interpreted electron densities and atomic models. E.B. and J.L. prepared figures. C.M.T.S. designed the study. E.B., J.L., and C.M.T.S. wrote the manuscript. All authors discussed the results and commented on the manuscript.

ACKNOWLEDGMENTS

We thank Helena Seibel and Brian Bauer for technical support during sample preparation and Alina Bretfeld and Mathias Br  nner for initial structural work. The present work was supported by grants from the Deutsche Forschungsgemeinschaft DFG (SFB 740 to C.M.T.S., P.S., and T.M.; SFB 1078 to P.S.), DFG Cluster of Excellence “Unifying Concepts in Catalysis” (Research Field D3/E3-1) to P.S., HSFP and Senatsverwaltung f  r Wissenschaft, Forschung und Kultur Berlin (UltraStructureNetwork, Anwerbezentrums), Charit   (Rahel-Hirsch stipend to T.V.B.), and by the NIH (R01 GM60635) to P.A.P. E.B. holds

a Freigeist-Fellowship from the Volkswagen Foundation. The authors acknowledge the North-German Supercomputing Alliance (HLRN) and the Texas Advanced Computing Center (TACC) at the University of Texas at Austin for providing high-performance computing resources that have contributed to the research results reported in this paper.

Received: September 2, 2014

Revised: January 5, 2015

Accepted: February 27, 2015

Published: May 7, 2015

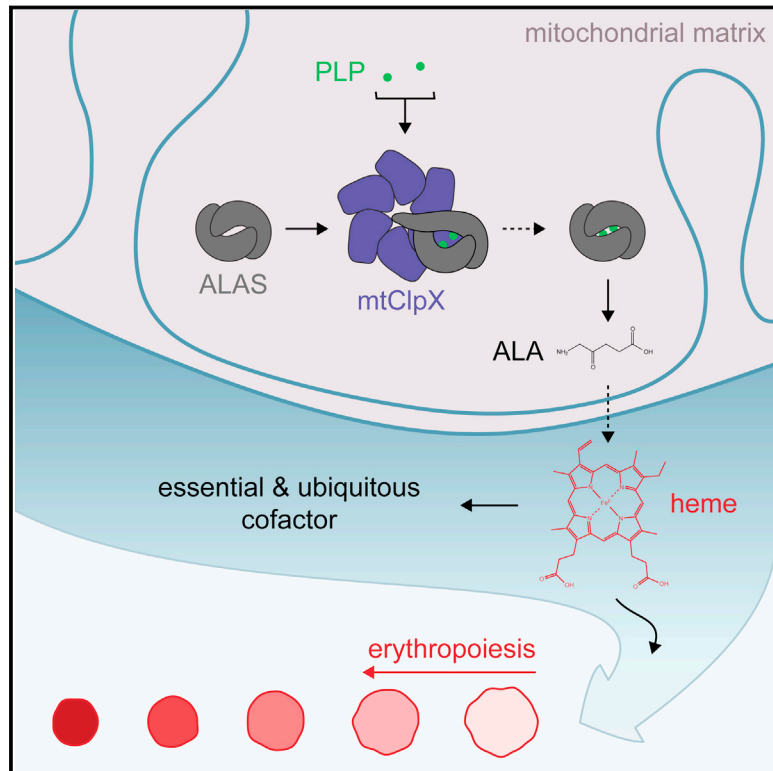
REFERENCES

- Agirrezabala, X., Lei, J., Brunelle, J.L., Ortiz-Meoz, R.F., Green, R., and Frank, J. (2008). Visualization of the hybrid state of tRNA binding promoted by spontaneous ratcheting of the ribosome. *Mol. Cell* 32, 190–197.
- Agmon, I., Bashan, A., Zarivach, R., and Yonath, A. (2005). Symmetry at the active site of the ribosome: structural and functional implications. *Biol. Chem.* 386, 833–844.
- Anger, A.M., Armache, J.-P., Berninghausen, O., Habeck, M., Subklewe, M., Wilson, D.N., and Beckmann, R. (2013). Structures of the human and *Drosophila* 80S ribosome. *Nature* 497, 80–85.
- Ben-Shem, A., Garreau de Loubresse, N., Melnikov, S., Jenner, L., Yusupova, G., and Yusupov, M. (2011). The structure of the eukaryotic ribosome at 3.0 Å resolution. *Science* 334, 1524–1529.
- Brandt, F., Carlson, L.-A., Hartl, F.U., Baumeister, W., and Grünwald, K. (2010). The three-dimensional organization of polyribosomes in intact human cells. *Mol. Cell* 39, 560–569.
- Budkevich, T., Giesebrecht, J., Altman, R.B., Munro, J.B., Mielke, T., Nierhaus, K.H., Blanchard, S.C., and Spahn, C.M.T. (2011). Structure and dynamics of the mammalian ribosomal pretranslocation complex. *Mol. Cell* 44, 214–224.
- Budkevich, T.V., Giesebrecht, J., Behrmann, E., Loerke, J., Ramrath, D.J.F., Mielke, T., Ismer, J., Hildebrand, P.W., Tung, C.-S., Nierhaus, K.H., et al. (2014). Regulation of the mammalian elongation cycle by subunit rolling: a eukaryotic-specific ribosome rearrangement. *Cell* 158, 121–131.
- Burnett, B.J., Altman, R.B., Ferrao, R., Alejo, J.L., Kaur, N., Kanji, J., and Blanchard, S.C. (2013). Elongation factor Ts directly facilitates the formation and disassembly of the *Escherichia coli* elongation factor Tu·GTP·aminoacyl-tRNA ternary complex. *J. Biol. Chem.* 288, 13917–13928.
- Burnett, B.J., Altman, R.B., Ferguson, A., Wasserman, M.R., Zhou, Z., and Blanchard, S.C. (2014). Direct evidence of an elongation factor-Tu/Ts·GTP·Aminoacyl-tRNA quaternary complex. *J. Biol. Chem.* 289, 23917–23927.
- Caulfield, T., and Devkota, B. (2012). Motion of transfer RNA from the A/T state into the A-site using docking and simulations. *Proteins* 80, 2489–2500.
- Demeshkina, N., Jenner, L., Westhof, E., Yusupov, M., and Yusupova, G. (2012). A new understanding of the decoding principle on the ribosome. *Nature* 484, 256–259.
- des Georges, A., Hashem, Y., Unbehaun, A., Grassucci, R.A., Taylor, D., Hellen, C.U.T., Pestova, T.V., and Frank, J. (2014). Structure of the mammalian ribosomal pre-termination complex associated with eRF1·eRF3·GDPNP. *Nucleic Acids Res.* 42, 3409–3418.
- Dunkle, J.A., and Cate, J.H.D. (2010). Ribosome structure and dynamics during translocation and termination. *Annu. Rev. Biophys.* 39, 227–244.
- Dyson, H.J. (2011). Expanding the proteome: disordered and alternatively folded proteins. *Q. Rev. Biophys.* 44, 467–518.
- Fischer, N., Konevega, A.L., Wintermeyer, W., Rodnina, M.V., and Stark, H. (2010). Ribosome dynamics and tRNA movement by time-resolved electron cryomicroscopy. *Nature* 466, 329–333.
- Frank, J. (2013). Story in a sample—the potential (and limitations) of cryo-electron microscopy applied to molecular machines. *Biopolymers* 99, 832–836.
- Frank, J., and Spahn, C.M.T. (2006). The ribosome and the mechanism of protein synthesis. *Rep. Prog. Phys.* 69, 1383–1417.
- Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., and Leith, A. (1996). SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J. Struct. Biol.* 116, 190–199.
- Frauenfelder, H., Sligar, S.G., and Wolynes, P.G. (1991). The energy landscapes and motions of proteins. *Science* 254, 1598–1603.
- Gabashvili, I.S., Agrawal, R.K., Spahn, C.M., Grassucci, R.A., Svergun, D.I., Frank, J., and Penczek, P. (2000). Solution structure of the *E. coli* 70S ribosome at 11.5 Å resolution. *Cell* 100, 537–549.
- Geggier, P., Dave, R., Feldman, M.B., Terry, D.S., Altman, R.B., Munro, J.B., and Blanchard, S.C. (2010). Conformational sampling of aminoacyl-tRNA during selection on the bacterial ribosome. *J. Mol. Biol.* 399, 576–595.
- Giegé, R., Jühling, F., Pütz, J., Stadler, P., Sauter, C., and Florentz, C. (2012). Structure of transfer RNAs: similarity and variability. *Wiley Interdiscip. Rev. RNA* 3, 37–61.
- Guo, Z., and Noller, H.F. (2012). Rotation of the head of the 30S ribosomal subunit during mRNA translocation. *Proc. Natl. Acad. Sci. USA* 109, 20391–20394.
- Hirashima, A., and Kaji, A. (1970). Factor dependent breakdown of polysomes. *Biochem. Biophys. Res. Commun.* 41, 877–883.
- Hopfield, J.J. (1974). Kinetic proofreading: a new mechanism for reducing errors in biosynthetic processes requiring high specificity. *Proc. Natl. Acad. Sci. USA* 71, 4135–4139.
- Jenner, L., Demeshkina, N., Yusupova, G., and Yusupov, M. (2010). Structural rearrangements of the ribosome at the tRNA proofreading step. *Nat. Struct. Mol. Biol.* 17, 1072–1078.
- Julián, P., Konevega, A.L., Scheres, S.H.W., Lázaro, M., Gil, D., Wintermeyer, W., Rodnina, M.V., and Valle, M. (2008). Structure of ratcheted ribosomes with tRNAs in hybrid states. *Proc. Natl. Acad. Sci. USA* 105, 16924–16927.
- Klein, D.J., Moore, P.B., and Steitz, T.A. (2004). The contribution of metal ions to the structural stability of the large ribosomal subunit. *RNA* 10, 1366–1379.
- Klinge, S., Voigts-Hoffmann, F., Leibundgut, M., Arpagaus, S., and Ban, N. (2011). Crystal structure of the eukaryotic 60S ribosomal subunit in complex with initiation factor 6. *Science* 334, 941–948.
- Korostelev, A., Ermolenko, D.N., and Noller, H.F. (2008). Structural dynamics of the ribosome. *Curr. Opin. Chem. Biol.* 12, 674–683.
- Loerke, J., Giesebrecht, J., and Spahn, C.M.T. (2010). Multiparticle cryo-EM of ribosomes. *Methods Enzymol.* 483, 161–177.
- Melnikov, S., Ben-Shem, A., Garreau de Loubresse, N., Jenner, L., Yusupova, G., and Yusupov, M. (2012). One core, two shells: bacterial and eukaryotic ribosomes. *Nat. Struct. Mol. Biol.* 19, 560–567.
- Moore, P.B. (2012). How should we think about the ribosome? *Annu. Rev. Biophys.* 41, 1–19.
- Munro, J.B., Sanbonmatsu, K.Y., Spahn, C.M.T., and Blanchard, S.C. (2009). Navigating the ribosome's metastable energy landscape. *Trends Biochem. Sci.* 34, 390–400.
- Myasnikov, A.G., Afonina, Z.A., Ménétret, J.-F., Shirokov, V.A., Spirin, A.S., and Klaholz, B.P. (2014). The molecular structure of the left-handed supra-molecular helix of eukaryotic polyribosomes. *Nat. Commun.* 5, 5294.
- Ogle, J.M., Carter, A.P., and Ramakrishnan, V. (2003). Insights into the decoding mechanism from recent ribosome structures. *Trends Biochem. Sci.* 28, 259–266.
- Pape, T., Wintermeyer, W., and Rodnina, M.V. (1998). Complete kinetic mechanism of elongation factor Tu-dependent binding of aminoacyl-tRNA to the A site of the *E. coli* ribosome. *EMBO J.* 17, 7490–7497.
- Pape, T., Wintermeyer, W., and Rodnina, M. (1999). Induced fit in initial selection and proofreading of aminoacyl-tRNA on the ribosome. *EMBO J.* 18, 3800–3807.
- Penczek, P.A., Fang, J., Li, X., Cheng, Y., Loerke, J., and Spahn, C.M.T. (2014). CTER-rapid estimation of CTF parameters with error assessment. *Ultramicroscopy* 140, 9–19.
- Petrov, A., Kornberg, G., O'Leary, S., Tsai, A., Uemura, S., and Puglisi, J.D. (2011). Dynamics of the translational machinery. *Curr. Opin. Struct. Biol.* 21, 137–145.

- Preis, A., Heuer, A., Barrio-Garcia, C., Hauser, A., Eyler, D.E., Berninghausen, O., Green, R., Becker, T., and Beckmann, R. (2014). Cryoelectron microscopic structures of eukaryotic translation termination complexes containing eRF1-eRF3 or eRF1-ABCE1. *Cell Rep.* 8, 59–65.
- Purcell, E.M. (1977). Life at low Reynolds number. *Am. J. Physiol.* 45, 3–11.
- Rabl, J., Leibundgut, M., Ataïde, S.F., Haag, A., and Ban, N. (2011). Crystal structure of the eukaryotic 40S ribosomal subunit in complex with initiation factor 1. *Science* 331, 730–736.
- Ratje, A.H., Loerke, J., Mikolajka, A., Br  nner, M., Hildebrand, P.W., Starosta, A.L., D  nh  fer, A., Connell, S.R., Fucini, P., Mielke, T., et al. (2010). Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid sites. *Nature* 468, 713–716.
- Rhodin, M.H.J., and Dinman, J.D. (2010). A flexible loop in yeast ribosomal protein L11 coordinates P-site tRNA binding. *Nucleic Acids Res.* 38, 8377–8389.
- Rich, A., Warner, J.R., and Goodman, H.M. (1963). The structure and function of polyribosomes. *Cold Spring Harb. Symp. Quant. Biol.* 28, 269–285.
- Sanbonmatsu, K.Y. (2006). Energy landscape of the ribosomal decoding center. *Biochimie* 88, 1053–1059.
- Schmeing, T.M., Moore, P.B., and Steitz, T.A. (2003). Structures of deacylated tRNA mimics bound to the E site of the large ribosomal subunit. *RNA* 9, 1345–1352.
- Schmeing, T.M., Voorhees, R.M., Kelley, A.C., Gao, Y.-G., Murphy, F.V., 4th, Weir, J.R., and Ramakrishnan, V. (2009). The crystal structure of the ribosome bound to EF-Tu and aminoacyl-tRNA. *Science* 326, 688–694.
- Schuette, J.-C., Murphy, F.V., 4th, Kelley, A.C., Weir, J.R., Giesebrecht, J., Connell, S.R., Loerke, J., Mielke, T., Zhang, W., Penczek, P.A., et al. (2009). GTPase activation of elongation factor EF-Tu by the ribosome during decoding. *EMBO J.* 28, 755–765.
- Selmer, M., Dunham, C.M., Murphy, F.V., 4th, Weixlbaumer, A., Petry, S., Kelley, A.C., Weir, J.R., and Ramakrishnan, V. (2006). Structure of the 70S ribosome complexed with mRNA and tRNA. *Science* 313, 1935–1942.
- Spahn, C.M.T., and Penczek, P.A. (2009). Exploring conformational modes of macromolecular assemblies by multiparticle cryo-EM. *Curr. Opin. Struct. Biol.* 19, 623–631.
- Spahn, C.M.T., Gomez-Lorenzo, M.G., Grassucci, R.A., J  rgensen, R., Andersen, G.R., Beckmann, R., Penczek, P.A., Ballesta, J.P.G., and Frank, J. (2004). Domain movements of elongation factor eEF2 and the eukaryotic 80S ribosome facilitate tRNA translocation. *EMBO J.* 23, 1008–1019.
- Stephens, S.B., and Nicchitta, C.V. (2007). In vitro and tissue culture methods for analysis of translation initiation on the endoplasmic reticulum. *Methods Enzymol.* 431, 47–60.
- Villa, E., Sengupta, J., Trabuco, L.G., LeBarron, J., Baxter, W.T., Shaikh, T.R., Grassucci, R.A., Nissen, P., Ehrenberg, M., Schulten, K., and Frank, J. (2009). Ribosome-induced changes in elongation factor Tu conformation control GTP hydrolysis. *Proc. Natl. Acad. Sci. USA* 106, 1063–1068.
- Voorhees, R.M., and Ramakrishnan, V. (2013). Structural basis of the translational elongation cycle. *Annu. Rev. Biochem.* 82, 203–236.
- Voorhees, R.M., Schmeing, T.M., Kelley, A.C., and Ramakrishnan, V. (2010). The mechanism for activation of GTP hydrolysis on the ribosome. *Science* 330, 835–838.
- Voorhees, R.M., Fern  ndez, I.S., Scheres, S.H.W., and Hegde, R.S. (2014). Structure of the mammalian ribosome-Sec61 complex to 3.4   resolution. *Cell* 157, 1632–1643.
- Whitford, P.C., Geggier, P., Altman, R.B., Blanchard, S.C., Onuchic, J.N., and Sanbonmatsu, K.Y. (2010). Accommodation of aminoacyl-tRNA into the ribosome involves reversible excursions along multiple pathways. *RNA* 16, 1196–1204.
- Wilson, D.N., and Nierhaus, K.H. (2006). The E-site story: the importance of maintaining two tRNAs on the ribosome during protein synthesis. *Cell. Mol. Life Sci.* 63, 2725–2737.

Mitochondrial ClpX Activates a Key Enzyme for Heme Biosynthesis and Erythropoiesis

Graphical Abstract



Authors

Julia R. Kardon, Yvette Y. Yien, ..., Barry H. Paw, Tania A. Baker

Correspondence

tabaker@mit.edu

In Brief

The ClpX unfoldase regulates bacterial proteomes largely by protein degradation. Mitochondria, which inherited chaperone proteins from their bacterial ancestor, have co-opted ClpX to stimulate cofactor incorporation into an essential heme biosynthesis enzyme.

Highlights

- The mitochondrial ClpX unfoldase is required for efficient heme biosynthesis
- mtClpX activates a key enzyme in heme biosynthesis by catalyzing cofactor binding
- mtClpX activates ALAS without committing it to degradation by mtClpP
- mtClpX is important for erythropoiesis when demand for heme is high



Mitochondrial ClpX Activates a Key Enzyme for Heme Biosynthesis and Erythropoiesis

Julia R. Kardon,^{1,7} Yvette Y. Yien,² Nicholas C. Huston,² Diana S. Branco,² Gordon J. Hildick-Smith,^{2,8} Kyu Y. Rhee,^{3,4} Barry H. Paw,^{2,5,6} and Tania A. Baker^{1,7,*}

¹Department of Biology, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

²Division of Hematology, Department of Medicine, Brigham and Women's Hospital, Harvard Medical School, Boston, MA 02115, USA

³Department of Microbiology and Immunology, Weill Cornell Medical College, New York, NY 10065, USA

⁴Division of Infectious Diseases, Department of Medicine, Weill Cornell Medical College, New York, NY 10065, USA

⁵Division of Hematology-Oncology, Department of Medicine, Boston Children's Hospital, Harvard Medical School, Boston, MA 02115, USA

⁶Department of Pediatric Oncology, Dana-Farber Cancer Institute, Boston, MA 02115, USA

⁷Howard Hughes Medical Institute, Massachusetts Institute of Technology, Cambridge, MA 02139, USA

⁸Present address: Weill Cornell Medical College, New York, NY 10065, USA

*Correspondence: tabaker@mit.edu

<http://dx.doi.org/10.1016/j.cell.2015.04.017>

SUMMARY

The mitochondrion maintains and regulates its proteome with chaperones primarily inherited from its bacterial endosymbiont ancestor. Among these chaperones is the AAA+ unfoldase ClpX, an important regulator of prokaryotic physiology with poorly defined function in the eukaryotic mitochondrion. We observed phenotypic similarity in *S. cerevisiae* genetic interaction data between mitochondrial ClpX (mtClpX) and genes contributing to heme biosynthesis, an essential mitochondrial function. Metabolomic analysis revealed that 5-aminolevulinic acid (ALA), the first heme precursor, is 5-fold reduced in yeast lacking mtClpX activity and that total heme is reduced by half. mtClpX directly stimulates ALA synthase in vitro by catalyzing incorporation of its cofactor, pyridoxal phosphate. This activity is conserved in mammalian homologs; additionally, mtClpX depletion impairs vertebrate erythropoiesis, which requires massive upregulation of heme biosynthesis to supply hemoglobin. mtClpX, therefore, is a widely conserved stimulator of an essential biosynthetic pathway and uses a previously unrecognized mechanism for AAA+ unfoldases.

INTRODUCTION

All organisms require AAA+ protein unfoldases to actively unfold selected proteins for protein quality control and to regulate the activity of specific substrates. The prokaryotic AAA+ unfoldase ClpX is particularly specialized for regulatory unfolding, tuning the proteome to respond to environmental stress and to orchestrate changes in cell state (Gottesman, 2003; Sauer et al., 2004). ClpX unfolds substrate proteins by ATP-driven translocation of the polypeptide chain through the central pore of its hexameric assembly. In complex with the ClpP peptidase, ClpX carries

out protein degradation by translocating unfolded substrates directly into the ClpP proteolytic chamber (Sauer et al., 2004). ClpP degrades all known substrates of ClpX, although for a few substrates, unfolding—not degradation—is the biologically required event (Konieczny and Helinski, 1997; Mhammedi-Alaoui et al., 1994).

In the eukaryotic cytoplasm, the 26S proteasome, which retains the basic architecture of Clp family proteases as well as a related AAA+ unfoldase component, functionally replaces the Clp family proteases. The mitochondrion, however, maintains an autonomous machinery for proteome remodeling, including ClpX, that is largely conserved from its α -proteobacterial ancestor (Figure S1). Mitochondrial ClpX (mtClpX) does not contribute substantially to protein quality control (Rottgers et al., 2002; van Dyck et al., 1998), suggesting that it may act primarily to control the activities of its substrates by regulatory unfolding and degradation, similarly to its prokaryotic homologs. Mitochondrial ClpP (mtClpP) is not as widely conserved as mtClpX, and mtClpX in organisms without ClpP lacks the ClpP interaction motif (Figure S1), suggesting that mtClpX may execute a protease-independent function. The specific contributions of mtClpX to mitochondrial physiology, however, are not well understood. mtClpX is required to initiate the mitochondrial unfolded protein response (Haynes et al., 2010) and has been observed to affect mitochondrial nucleoid morphology (Bogenhagen et al., 2008; Kasashima et al., 2012), but its mechanism in these roles is unknown. The single physiological substrate identified for mtClpX, the GTPase Noa1, is degraded by mtClpXP, but how this degradation contributes to Noa1 maintenance or regulation in vivo is unclear (Al-Furoukh et al., 2014).

To uncover the physiological functions and partners of mtClpX, we mined previously generated large-scale genetic and chemical interaction maps in *S. cerevisiae* (Costanzo et al., 2010; Hoppins et al., 2011; Lee et al., 2014). We observed strong links between the yeast mtClpX gene (*MCX1*) and genes involved in the first steps of heme biosynthesis (Figure 1A), suggesting that mtClpX might act during heme biosynthesis as well.

Nearly all organisms (with a few known exceptions among parasites) require heme for viability (Kořený et al., 2012), and most organisms synthesize heme endogenously. Heme is an

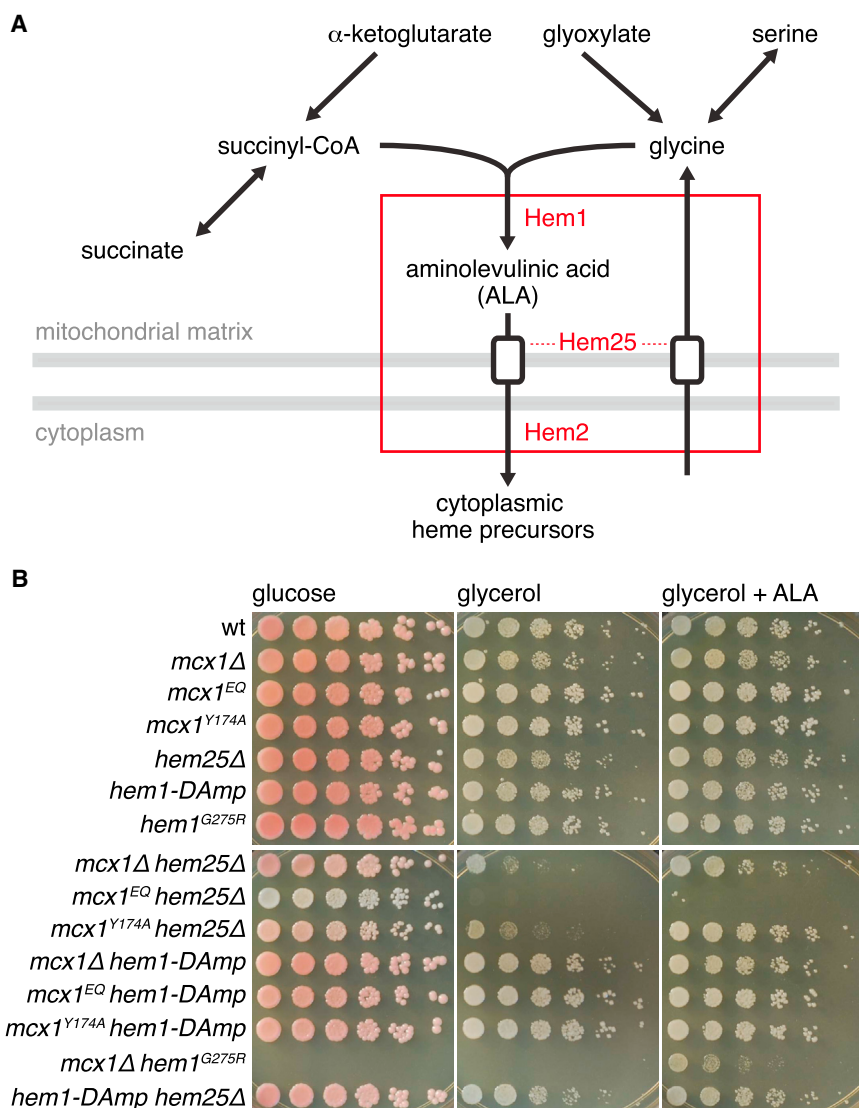


Figure 1. *MCX1* Interacts Chemically and Genetically with the Heme Biosynthetic Pathway

(A) The metabolic pathway for the first step of heme biosynthesis in non-plant eukaryotes. The genetic and chemical interaction profile of *MCX1* is highly correlated with the profiles of yeast genes (*HEM25*, *HEM1*, and *HEM2*, shown in red) involved in the first steps of heme biosynthesis. Dashed lines indicate uncertainty in assigning Hem25 to glycine uptake or ALA export. Gray bars indicate mitochondrial membranes. CoA, coenzyme A. (B) *MCX1*, *HEM1*, and *HEM25* alleles exhibit synthetic phenotypes. 5-fold serial dilutions from cell suspensions with optical density 600 (OD_{600}) = 1 were pinned on YP (1% yeast extract, 2% peptone) + 2% agar, with 2% glucose or 3% glycerol. ALA in “glycerol + ALA” indicates 50 μ g/ml ALA. Growth on glucose after 2 days and on glycerol after 3 days is shown. wt, wild-type. See also Figure S1.

essential cofactor for many enzymes, including several members of the respiratory chain, p450 enzymes, and sterol biosynthetic enzymes, and also acts as the sensor component of multiple environmentally responsive transcription factors (Girvan and Munro, 2013; Hamza and Dailey, 2012). In non-plant eukaryotes, the first, rate-limiting step of heme biosynthesis is carried out in the mitochondrial matrix, and its product, 5-aminolevulinic acid (ALA), is exported to the cytoplasm (Figure 1A). After several further biosynthetic steps, a heme precursor is re-imported into the mitochondrion, where synthesis is completed (Hamza and Dailey, 2012). Cells tightly control heme biosynthesis to meet demand; overstimulation of heme biosynthesis drains valuable central metabolites and can cause damage from reactive unliganded heme or accumulation of toxic heme precursors, whereas insufficient heme production limits the activity of the diverse proteins that require it as a cofactor (Girvan and Munro, 2013; Hamza and Dailey, 2012). As a consequence, causative human disease alleles of every enzyme in heme

biosynthesis have been identified (Cama-schella, 2009; Sassa, 2006).

In this study, we discover a stimulatory function for mtClpX in heme biosynthesis. Comparison of metabolite levels in wild-type and *mcx1Δ* cell extracts indicated that *MCX1* acts to promote ALA synthesis, the initial step of heme biosynthesis. *Mcx1* directly activates the enzyme that performs this step, ALA synthase (ALAS, Hem1 in yeast), by accelerating binding of the cofactor pyridoxal phosphate (PLP) to apoenzyme. This activation is conserved for mammalian homologs of these enzymes and proceeds without degradation by mtClpP. mtClpX, therefore, stimulates an essential biosynthetic process through a previously unrecognized activity for AAA+ unfoldases: accel-

erating cofactor insertion into its protein substrate. Finally, we find that vertebrate erythropoiesis is impaired by mtClpX knock-down, commensurate with a central, conserved role for mtClpX in heme production.

RESULTS

MCX1 Promotes Heme Biosynthesis

Using *S. cerevisiae* genetic (Costanzo et al., 2010; Hoppins et al., 2011) and chemical (Lee et al., 2014) interaction data, we searched for genes with interaction profiles similar to that of *MCX1*, the yeast gene encoding mtClpX (van Dyck et al., 1998). Because *S. cerevisiae* (like several other fungi), lacks a ClpP homolog, yeast *Mcx1* likely acts purely as a protein unfoldase, without coupled degradation of its substrates. *MCX1* was strongly correlated (Costanzo et al., 2010; Hoppins et al., 2011; Lee et al., 2014) with several genes involved in the early steps in heme biosynthesis: *HEM1*, the gene encoding ALAS

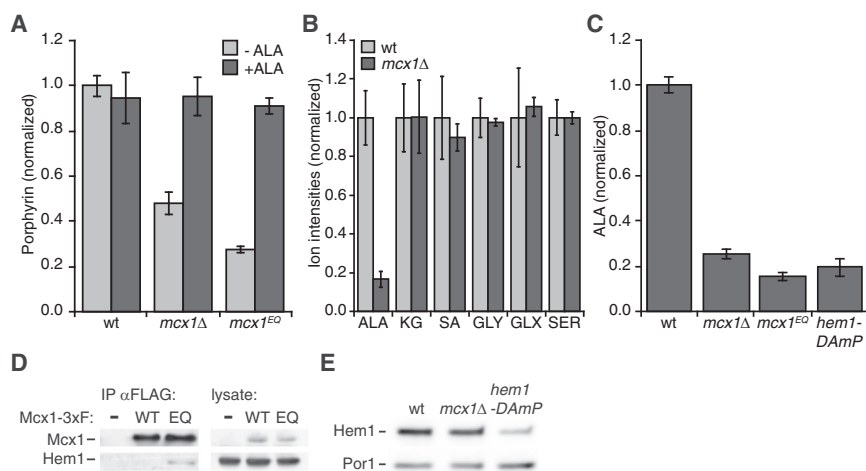


Figure 2. Mcx1 Promotes Heme Biosynthesis at the Step of ALA Synthesis

(A) Total porphyrin levels were measured by fluorescence in oxalic acid cell extracts (excitation, 400 nm; emission, 662 nm); $p < 0.001$ for difference between wild (wt) and *MCX1* mutants. +ALA indicates supplementation of growth medium with 50 $\mu\text{g/ml}$ ALA.

(B) Metabolites involved in the first step of heme biosynthesis (KG, α -ketoglutarate; SA, succinic acid; GLY, glycine; GLX, glyoxylate; SER, serine) were measured in extracts of the indicated yeast strains by LC-MS. $p < 0.001$ for ALA perturbation in *mcx1Δ* cells.

(C) ALA levels in cell extracts were measured using modified Ehrlich's reagent. $p \leq 10^{-5}$ for ALA reduction in *MCX1* and *HEM1* mutants.

(D) Mcx1 was isolated with α -FLAG antibody-conjugated beads from cells harboring *HEM1*-3xMYC and *MCX1*-3xFLAG (wild-type [WT] or

Mcx1^{EQ} [EQ]) or untagged *Mcx1* [–] alleles at the genomic loci and eluted with 3xFLAG peptide. The eluate was analyzed by western blot for Mcx1 (α -FLAG) and Hem1 (α -Myc). The image below the label “IP α -FLAG” represents the proteins that were immunoprecipitated with an anti-FLAG antibody.

(E) Cellular levels of Hem1-3xMyc were analyzed by western blot, using alkaline cell extracts (von der Haar, 2007). Hem1-3xMyc intensity: in *mcx1Δ*, 1.1 ± 0.1 , relative to wild-type, $p = 0.35$ for difference; in *hem1-DAmP*, 0.3 ± 0.1 , $p = 0.01$. The mitochondrial protein Por1 was probed as a loading control.

Error bars represent mean \pm SD. See also Extended Experimental Procedures and Figure S2.

(EC 2.3.1.37) (Arrese et al., 1983); *HEM25*, encoding the putative mitochondrial glycine/ALA transporter SLC25A38 (Guernsey et al., 2009); and *HEM2*, encoding the cytosolic enzyme that catalyzes the second step in heme biosynthesis, ALA dehydratase (ALAD; EC 4.2.1.24) (Gollub et al., 1977; Figure 1A).

We tested the growth of *MCX1*, *HEM1*, and *HEM25* mutants singly and in combination on fermentable (glucose) or mitochondrial respiration-requiring (glycerol) carbon sources (Figure 1B). *MCX1* deletion or mutation to an ATP hydrolysis-blocked allele (mutation E206Q in the Walker B motif, *mcx1*^{EQ}) did not impair fermentative or respiratory growth. A mutation in the essential gene *HEM1* corresponding to a sideroblastic anemia allele of human ALAS (G351R in human ALAS2 [Wintrobe and Greer, 2004], G275R in yeast *HEM1*) also showed normal fermentative and respiratory growth but was lethal in combination with *mcx1Δ* (Figure 1B). In combination with the deletion of *HEM25*, *MCX1* deletion or mutation dramatically impaired or abrogated mitochondrial respiration (Figure 1B). Respiratory growth of *mcx1Δ hem25Δ*, and *mcx1Δ hem1*^{G275R} was restored by supplementation with ALA, indicating that the synthetic phenotypes of Mcx1 resulted from a deficiency at the mitochondrial first step in heme biosynthesis. These strong synthetic phenotypes suggest an important role for Mcx1 in heme biosynthesis.

To directly test the contribution of *MCX1* to heme biosynthesis, we measured heme levels in logarithmically growing yeast by total porphyrin fluorescence and by ⁵⁵Fe incorporation. Both measurements indicated a 2- and a 3-fold reduction of heme in *mcx1Δ* and *mcx1*^{EQ} yeast, respectively (Figure 2A; Figure S2A). Mirroring the severity of its respiratory growth phenotype, *mcx1Δ hem25Δ* yeast (Figure S2B) exhibited a greater heme deficiency than *MCX1* mutant yeast. Supplementing the growth medium with ALA rescued the poor heme production of *MCX1* mutants (Figure 2A). This rescue strongly suggests that Mcx1 promotes the first phase of heme biosynthesis: synthesis and export of ALA.

Mcx1 Stimulates ALA Synthesis, the First Step of Heme Biosynthesis

To determine the specific perturbation leading to heme deficiency in cells lacking *MCX1*, we monitored total metabolites in extracts from wild-type and *mcx1Δ* yeast by liquid chromatography-mass spectrometry (LC-MS). Because *MCX1*-correlated genes function early in heme biosynthesis, we focused on metabolites directly involved in the first mitochondrial phase (Figure 1A). Ion intensities detected for metabolites preceding ALA synthesis (α -ketoglutarate, succinate, glycine, glyoxylate, and serine) were equivalent in wild-type and *mcx1Δ* extracts. In contrast, the ion intensity for ALA was reduced more than 80% in *mcx1Δ* extracts (Figure 2B). Chemical detection of ALA in extracts corroborated this reduction (Figure 2C). *mcx1Δ* extracts exhibited 75% reduced ALA in this assay, and *mcx1*^{EQ} extracts had an enhanced defect (~85% reduction) (Figure 2C). For comparison, a reduced expression allele of Hem1 (*hem1-DAmP*) (Schuldiner et al., 2005) caused an 80% reduction in cellular ALA. Thus, we conclude that Mcx1 activity promotes heme production at the first step of ALA synthesis.

Mcx1 Interacts Directly with the ALAS Hem1

We hypothesized that the Mcx1 unfoldase could act directly on Hem1 (the yeast ALAS) to promote ALA synthesis. ALAS is the rate-limiting enzyme for heme biosynthesis in nearly all cell types (Hamza and Dailey, 2012) and, as such, would be a likely target for a stimulatory factor in heme biosynthesis. To test for physical interaction between Mcx1 and Hem1, we affinity purified Mcx1-3xFLAG and Mcx1^{EQ}-3xFLAG from yeast cell extracts and probed for copurifying Hem1-3xMyc (Figure 2D). We detected Hem1 in Mcx1^{EQ} but not wild-type Mcx1 purified samples, consistent with the ATP dependence of ClpX-substrate interactions. These data suggest that Hem1 is a direct substrate for the Mcx1 unfoldase.

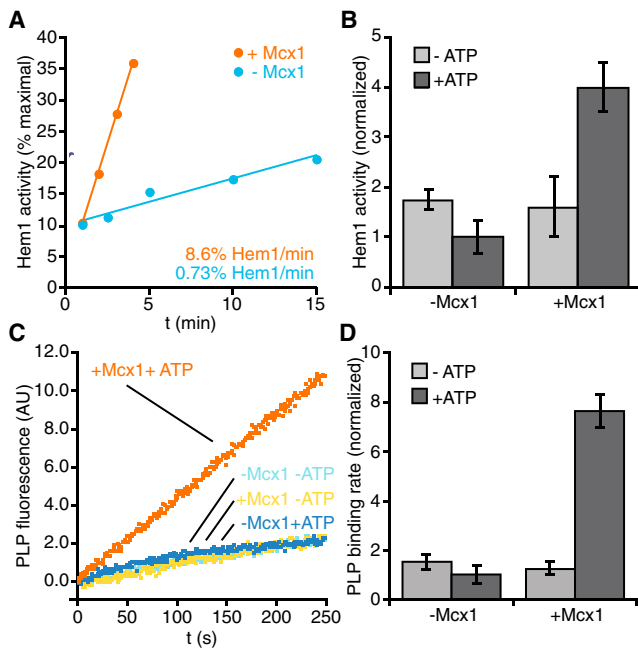


Figure 3. Mcx1 Accelerates Incorporation of PLP Cofactor into Hem1

(A) Rate of apoHem1 activation by PLP. apoHem1 (3 μ M) was incubated with PLP (50 μ M) and ATP (2 mM), with (orange) or without Mcx1 (2 μ M hexamer) (blue), and assayed for ALAS activity at indicated times using modified Ehrlich's reagent. t, time.

(B) ALAS activity resulting from 4 min incubation of apoHem1 \pm Mcx1, with or without ATP, assayed as in (A). $p < 0.0001$ for stimulation by Mcx1 + ATP.

(C) PLP binding to apoHem1 was monitored by pyridoxyllysine fluorescence (excitation, 434 nm; emission, 515 nm).

(D) Rates of PLP binding to apoHem1 determined by linear fits to fluorescence increase between 100 s and 200 s in (C). $p < 0.0001$ for stimulation by Mcx1. Error bars represent mean \pm SD.

See also Figure S3.

To increase ALA production by acting on Hem1, Mcx1 might increase Hem1 protein abundance or might stimulate the activity of Hem1. Multiple mechanisms for control of ALAS abundance have been characterized (Hamza and Dailey, 2012; Tian et al., 2011). If Mcx1 is required to maintain Hem1 protein levels, decreased Hem1 would be expected in *mcx1* Δ cells. Hem1 protein abundance was equivalent in wild-type and *mcx1* Δ lysates (Figure 2E). In contrast, Hem1 reduction in a *hem1-DAmP* strain was easily detected (Figure 2E).

To test whether Mcx1 acts by increasing the enzymatic activity of Hem1, we purified these proteins and tested the effect of Mcx1 on the catalytic activity of Hem1. In contrast to the large reduction in ALA in *mcx1* Δ cells, Mcx1 had little effect on the rate of ALA production by Hem1 in vitro (Figure S3), leading us to consider mechanisms by which Mcx1 might stimulate Hem1 activity other than a straightforward effect on V_{\max} .

Mcx1 Activates Hem1 by Stimulating Insertion of Its Cofactor

Hem1 (like all ALASs) is part of a large, evolutionarily related enzyme family that depends on the cofactor PLP for activity

(the α -family of PLP-dependent enzymes) (Eliot and Kirsch, 2004). PLP binds covalently to an active-site lysine but requires other contacts buried in the interface of the homodimeric enzyme for stable binding (Astner et al., 2005; Gong et al., 1996). We hypothesized that Mcx1 might stimulate formation of the PLP-loaded holoenzyme. To test this idea, we prepared PLP-free Hem1 (apoHem1) and monitored enzyme activity after the addition of PLP. As reported previously (Volland and Felix, 1984), apoHem1 regained activity slowly on its own (0.73%/min) (Figure 3A). Inclusion of Mcx1 accelerated apoHem1 activation by a factor of ten (Figure 3A). Acceleration by Mcx1 depended on ATP, as expected if Mcx1 acts by remodeling or unfolding Hem1 (Figure 3B). We directly measured PLP binding to apoHem1 by monitoring the formation of the fluorescent pyridoxyllysine bond. The rate of PLP binding for apoHem1 alone (1.0%/min) (Figure 3C) closely matched the rate of activation for apoHem1 alone (Figure 3A). Mcx1 stimulated PLP binding by a factor of eight (7.6%/min), and this stimulation was also dependent on ATP (Figures 3C and 3D). Thus, Mcx1 accelerates formation of active Hem1 by stimulating cofactor binding to the apoenzyme.

We expect that, to efficiently activate apoHem1 in mitochondria, Mcx1 would interact preferentially with apoHem1 over holoHem1. Using purified proteins, we monitored the interaction of Mcx1-3xFLAG with apo- and holoHem1 by coimmunoprecipitation. We observed more efficient interaction of purified Hem1 with ATP-locked Mcx1 (Mcx1^{EQ}) than with wild-type Mcx1 (Figures 4B and 4C), consistent with the interaction observed between Hem1 and ATP-locked Mcx1 in cell extracts. Notably, we observed a 10-fold greater amount of Mcx1-bound apoHem1 than holoHem1 (Figures 4B and 4C), indicating that Mcx1 has intrinsic binding specificity for the species of Hem1 that it stimulates.

Mcx1 Activates Hem1 by Acting as an Unfoldase

Most activities characterized for AAA+ unfoldases involve complete or large-scale unfolding of their substrates. Because the folded context of the ALAS active site is important for PLP binding, Mcx1 is unlikely to globally unfold the enzyme to activate it. Therefore, we sought to determine whether Mcx1 acts as an unfoldase to activate Hem1 or whether it might use an alternative mechanism. To unfold their substrates, AAA+ unfoldases translocate the substrate polypeptide through the hexamer pore by the ATP-powered movement of several loops within the pore (Figure 4A; Figure S1). To test whether activating Hem1 requires the unfolding machinery of Mcx1, we introduced mutations into these loops at the genomic locus. The central loop (pore-1) is essential for gripping and translocating polypeptides through the pore of the hexamer; we mutated the invariant tyrosine (*mcx1*^{Y174A}) within this highly conserved sequence, an alteration that abrogates the unfolding activity of many AAA+ unfoldases, including ClpX (Siddiqui et al., 2004). The RKH and pore-2 loops contribute to, but are less critical for, unfolding. They are more important for substrate selection, and their sequences diverge widely across evolution (Martin et al., 2008a) (Figure S1). We replaced the RKH or pore-2 loops with the highly divergent *E. coli* sequences (*mcx1*^{E.c.RKH} or *mcx1*^{E.c.pore-2}; Figure S1; Table S1), a substitution that was previously demonstrated to transplant

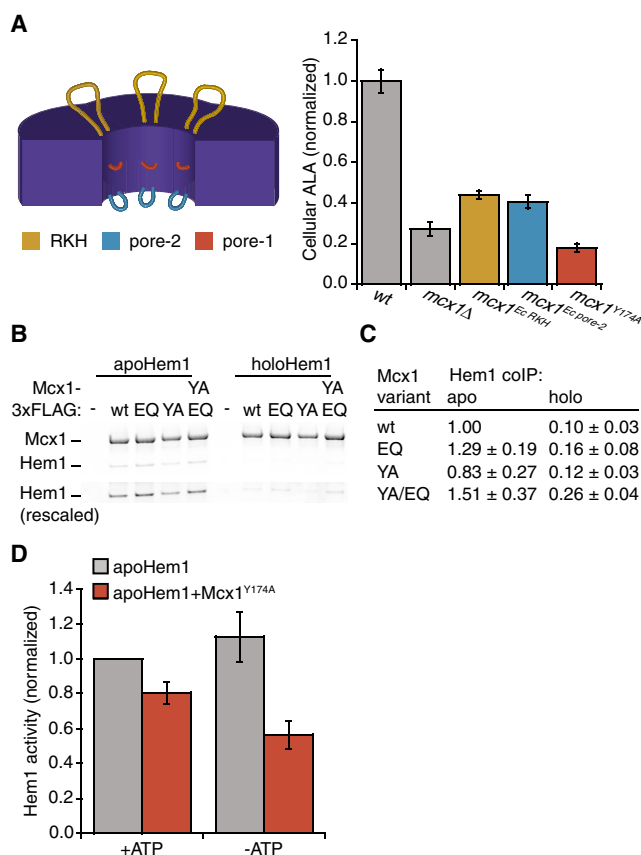


Figure 4. Mcx1 Requires the ClpX Translocating Pore Loops to Activate Hem1 and Promote ALA Production

(A) Left: pore loops are highlighted on a cross-section diagram of a ClpX hexamer. RKH loops are shown in yellow, pore-1 loops are in dark orange, and pore-2 loops are in blue. Right: ALA levels in cell extracts were measured by modified Ehrlich's reagent and normalized to wild-type (wt). Cells harbored indicated mutations at the genomic *MCX1* locus. $p < 0.0001$ for ALA reduction in all *MCX1* mutants.

(B) Co-immunoprecipitation of apo- and holoHem1 with Mcx1-3xFLAG variants was tested, using purified proteins. Proteins were separated by SDS-PAGE and stained with Sypro Orange. Lower panel, "Hem1 (rescaled)," shows Hem1 bands rescaled to maximum Hem1 intensity. Mcx1 variants are indicated as follows: wt, wild-type; EQ, Walker B E206Q; and YA, pore-1 Y174A.

(C) Quantitation of Hem1 intensities in (B), normalized to Mcx1-apoHem1 coimmunoprecipitation. $p < 0.001$ for more apoHem1 than holoHem1 bound by each Mcx1 variant. apo, apoHem1; holo, holoHem1.

(D) ALAS activity resulting from 4 min incubation of apoHem1 (apo) with 50 μM PLP ± Mcx1^{Y174A}, ± ATP, assayed as in Figure 3A. $p < 0.05$ for suppression of Hem1 by Mcx1^{Y174A} both with and without ATP. colP, co-immunoprecipitation. Error bars represent mean ± SD.

See also Figure S1.

E. coli ClpX substrate specificity to human mtClpX (Martin et al., 2008a). To determine the ability of these variant enzymes to activate Hem1, we monitored ALA levels within the corresponding *MCX1* mutant strains. All pore loop mutants had reduced ALA, and the magnitude of reduction correlated with the importance of the pore loop to translocation (Figure 4A). The pore-1 mutation caused a more severe ALA deficiency than *MCX1* deletion, and substitution with *E. coli* RKH or pore-2 loops caused a milder

ALA deficiency than *MCX1* deletion. *mcx1*^{Y174A} also exhibited genetic interactions similar to those of *mcx1Δ* (Figure 1B).

To further probe the contribution of the ClpX translocation machinery to ALA synthesis, we tested the activity of purified Mcx1^{Y174A} in vitro. As observed previously for *E. coli* ClpX (Martin et al., 2008b), Mcx1^{Y174A} has mildly elevated ATPase activity (368 ± 55 ATP per hexamer per minute, compared to 219 ± 23 ATP per hexamer per minute for wild-type), indicating that Mcx1^{Y174A} phenotypes do not result from loss of ATPase activity. Because Mcx1 pore loops might bind unstructured elements of Hem1, ALA reduction in Mcx1 pore loop mutants could be due to a defect in binding Hem1 rather than unfolding. To test Mcx1^{Y174A} substrate binding, we monitored the interaction of Hem1 with Mcx1^{Y174A} by coimmunoprecipitation using purified proteins. We observed that Hem1 interacted with equivalent strength to Mcx1^{Y174A} and wild-type Mcx1 and also was selective for the apoenzyme by an order of magnitude (Figures 4B and 4C). Therefore, mutation of the essential translocating ClpX pore loop-1 does not abolish Mcx1 stimulation of Hem1 by disrupting complex formation. Having excluded an interaction defect, we tested the effect of Mcx1^{Y174A} on PLP activation of apoHem1 in vitro. Mcx1^{Y174A} mildly suppressed apoHem1 activation (Figure 4D; compare to Figure 3B for stimulation by Mcx1), suggesting that nonproductive interactions of translocation-blocked Mcx1 with Hem1 interfere with Hem1 spontaneous activation. These data strongly indicate that Mcx1 activates Hem1 using the central polypeptide translocating activity of ClpX homologs.

mtClpX Activation of ALAS Is Conserved in Metazoans

Sequence conservation is high between fungal and metazoan ALAS and mtClpX, suggesting that activation of ALAS by mtClpX may be conserved. We measured the effect of mouse mtClpX on PLP activation of human erythroid ALAS (ALAS2) apoenzyme in vitro. mtClpX stimulated apoALAS2 activation approximately 2.5-fold (Figure 5A); this activation required the presence of ATP (Figure S4). mtClpX activation of ALAS, therefore, is broadly conserved among eukaryotes.

Although *S. cerevisiae* lacks a ClpP homolog, mtClpP is present in most other eukaryotes. Because the effect of mtClpX on ALAS is activating, the presence of a mtClpP protease might oppose this action of mtClpX. Therefore, we wanted to determine whether mtClpP interferes with this activation by coupled degradation. In the presence of mtClpP, mtClpX still stimulated PLP activation of apoALAS2, but the magnitude of this stimulation was reduced (Figure 5A). We monitored possible apoALAS2 degradation during incubation with mtClpXP and ATP. No degradation of apoALAS2 was observed over 30 min (Figures 5B and 5C), although mtClpXP efficiently degraded casein (a model substrate for mtClpXP) (Kang et al., 2002) under the same conditions.

We considered whether mtClpP might suppress mtClpX activation of apoALAS2 by suppressing mtClpX ATPase activity. Prokaryotic ClpP partially suppresses the ATPase of prokaryotic ClpX (Kim et al., 2001). If present in the mitochondrial enzymes, ATPase suppression could explain the lower stimulation of ALAS2 by mtClpX when mtClpP is present. In the presence of mtClpP, we observed an ~30%–40% suppression of the mtClpX

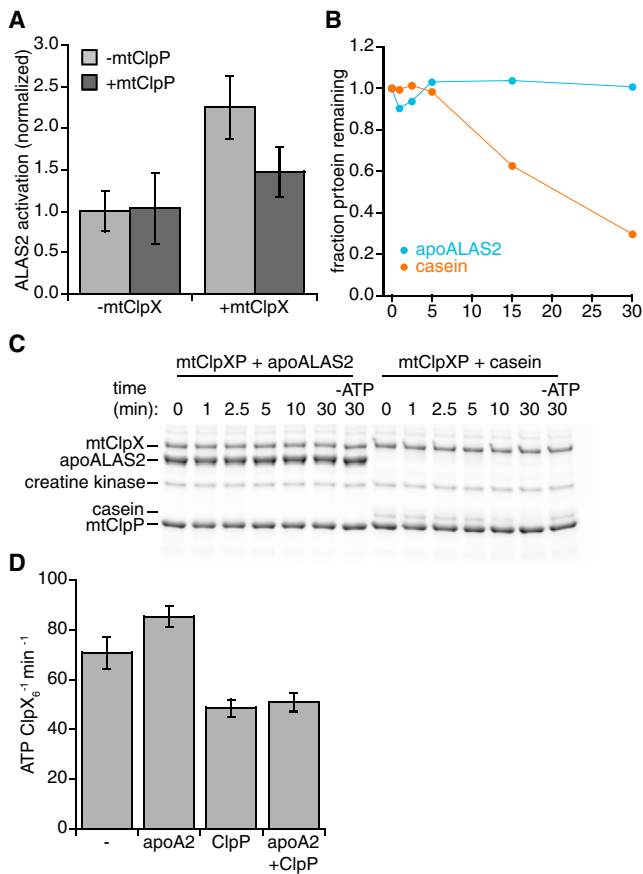


Figure 5. Mammalian mtClpX Stimulates PLP Activation of apoALAS2 and Does Not Direct apoALAS2 for Degradation by mtClpP

(A) apoALAS2 activation by PLP in vitro. Recombinant human apoALAS2 (5 μ M) was incubated with PLP, with mouse mtClpX (2 μ M hexamer) and human mtClpP (3 μ M 14-mer) as indicated, and activation between 4 and 10 min was measured by an NAD-coupled assay. $p < 0.01$ for acceleration of apoALAS2 activation by mtClpX, and $p < 0.05$ for acceleration by mtClpXP. (B and C) mtClpXP degradation test. apoALAS2 or α -casein (5 μ M each) were incubated with mouse mtClpX (0.3 μ M hexamer), human mtClpP (0.8 μ M 14-mer), and ATP regenerating system (including 4 mM ATP, except where noted) at 30°C, and aliquots were withdrawn and quenched with SDS at indicated time points. Proteins were separated by SDS-PAGE and stained with Sypro Orange. Quantitation of degradation is shown in (B), and gel is shown in (C). (D) Mouse mtClpX (0.3 μ M hexamer) ATPase activity was monitored by NADH-coupled assay, in the presence of human mtClpP (0.8 μ M 14-mer) and apoALAS2 (apoA2, 10 μ M) as indicated. $p < 0.01$ for suppression of ATPase by mtClpP, and $p < 0.05$ for stimulation of ATPase by apoALAS2. Error bars represent mean \pm SD. See also Figure S4.

ATPase rate (Figure 5D). This effect is commensurate with the magnitude of suppression of ALAS2 activation and could, therefore, account for reduced activation by mtClpXP without invoking protein degradation. We did not observe strong stimulation of mtClpX ATPase (either basal or ClpP suppressed) by ALAS2 (Figure 5D). Stimulation of ClpX ATPase by its substrates is common but highly variable and is less apparent for substrates that are more resistant to unfolding (Burton et al., 2003; Kenniston et al., 2003).

mtClpX Is Required for Efficient Erythropoiesis

In organisms with circulating blood cells, heme biosynthesis is massively upregulated during erythropoiesis to meet demand from hemoglobin production (Hamza and Dailey, 2012). Defects in heme biosynthesis cause several human anemias; congenital sideroblastic anemia is caused most commonly by mutations in ALAS2, the erythroid-specific ALA synthase (Camaschella, 2009). Because mtClpX activates ALAS, we reasoned that it would be crucial for erythropoiesis. To facilitate increased heme production during erythropoiesis, heme biosynthetic genes are transcriptionally upregulated. We examined a previously published genome-wide transcriptional dataset for human hematopoiesis (Novershtern et al., 2011) and observed that mRNA levels of CLPX, but not CLPP, were upregulated during erythropoiesis (Figure 6A). We also observed upregulation of *Clpx* during erythroid maturation in Friend mouse erythroleukemia (MEL) cells (Figure S5A). These data suggest that CLPX contributes to erythropoiesis. In addition, upregulation of CLPX, but not CLPP, suggests that in this cell type, increased CLPX relative to CLPP may help to avoid CLPP suppression of ALAS activation.

To examine the contribution of mtClpX to red blood cell development, we performed morpholino-mediated knockdowns in *D. rerio*. Zebrafish encode two homologs of mtClpX, *clpxa* and *clpxb*. Expression of both *clpxa* and *clpxb* is ubiquitous throughout the embryo (Figure S5B), indicating that mtClpX function is likely not restricted to developing red blood cells. *clpxa* mRNA was specifically reduced by two independent morpholino sequences (Figure S5C; Table S2); both morpholinos resulted in morphologically normal embryos with reduced hemoglobin staining by *o*-dianisidine, indicating a specific defect in red blood cell development (Figure 6B). Morpholinos targeting *clpxb* caused toxicity without specific mRNA reduction, preventing further phenotypic analysis. To quantify the anemia we observed in *clpxa* knockdown embryos, we performed flow cytometry of dissociated cells from embryos with GFP-marked erythroid cells; *clpxa* knockdown embryos from both morpholinos exhibited an ~50% reduction in GFP-positive cells at 72 hr post-fertilization (hpf) (Figure 6C). We also observed a reduction in early erythroid precursors at 24 hpf (Figure S5F), which may result from reduction in non-erythroid heme production from ALAS1 (Okano et al., 2010). ALAS1 shares very high sequence identity with the erythroid-specific ALAS2, as well as with yeast ALAS, and is likely subject to the same stimulatory activity by mtClpX. Supplementation with ALA starting at 24 hpf fully rescued *clpxa* knockdown-induced anemia (Figure 6D). ALA supplementation specifically rescued anemia in ALAS2 mutant embryos (*sauternes*; Brownlie et al., 1998), but not anemia in mitochondrial iron transporter (MFRN1, SLC25A37) mutants (*frascati*; Shaw et al., 2006) (Figure 6E). These results demonstrate that ALA specifically rescues defects in ALA synthesis and not defects in later steps in heme biosynthesis. Therefore, mtClpX stimulation of ALA synthesis is conserved from *S. cerevisiae* to vertebrates and is important for efficient heme synthesis during erythropoiesis.

DISCUSSION

In this work, we used large-scale genetic interaction data coupled with metabolic analysis to uncover a broadly conserved

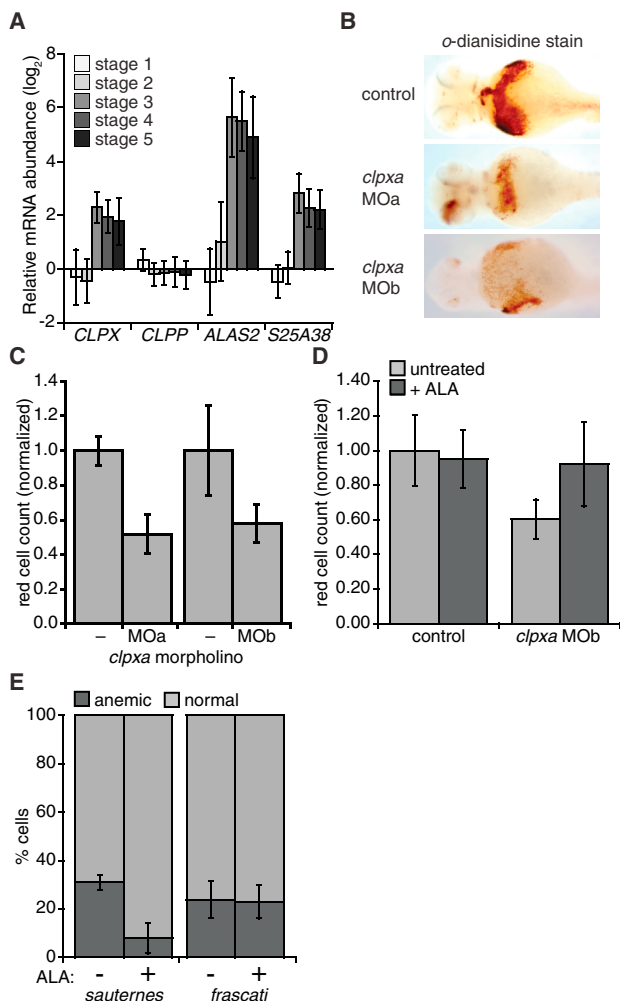


Figure 6. mtClpX Is Important for Vertebrate Heme Biosynthesis and Erythropoiesis

(A) Relative mRNA abundance for human *CLPX*, *CLPP*, *ALAS2*, and *SLC25A38* (indicated as *S25A38*) throughout erythropoiesis, as indicated in a microarray dataset described in Novershtern et al. (2011). Erythroid development stages were defined by cell-type-specific markers as follows: 1, CD34+ CD71+ GlyA-; 2, CD34- CD71+ GlyA-; 3, CD34- CD71+ GlyA+; 4, CD34- CD71low GlyA+; 5, CD34- CD71- GlyA+.

(B) o-dianisidine staining (brown) for hemoglobinized red cells in zebrafish embryos. Embryos were grown from zygotes injected at the one- to two-cell stage with *clpxa*-targeting morpholinos (MOa and MOb) or uninjected zygotes (control).

(C) Erythrocyte development at 72 hpf was quantified by flow cytometry, using dissociated cells from *Tg(globin-LCR:eGFP)* zebrafish. *p* ≤ 0.01 for erythrocyte reduction by *clpxa* knockdown with either morpholino.

(D) Rescue of *clpxa* MOb-induced anemia by ALA supplementation. *Tg(globin-LCR:eGFP)* zebrafish embryos were supplemented with 2 mM ALA from 24 to 72 hpf, upon which GFP+ erythrocytes were quantified by flow cytometry. *p* = 0.025 for rescue of anemia in *clpxa* knockdown embryos by ALA supplementation.

(E) Heterozygous *sauternes* (*ALAS2* mutant) or *frascati* (*SLC25A37* mutant) zebrafish were crossed, and progeny were grown for 72 hpf, with or without ALA supplementation as in (D). Anemia was assayed by o-dianisidine staining. *p* = 0.04 for rescue of anemia in *sauternes*^{+/-} progeny by ALA. *n* = 52 for *sauternes* -ALA; *n* = 43 for *sauternes* +ALA; *n* = 98 for *frascati* -ALA; *n* = 122 for *frascati* +ALA.

Error bars represent mean ± SD. See also Figure S5.

stimulatory activity of mtClpX in the essential biological process of heme biosynthesis. Our biochemical studies revealed that mtClpX specifically activates the apoenzyme form of ALAS, the first enzyme in heme biosynthesis, by accelerating binding to the cofactor PLP. mtClpX activation of ALAS is not coupled to degradation by mtClpP, although the presence of mtClpP results in partial inhibition. Consistent with its function in heme biosynthesis, mtClpX is important during erythropoiesis, when heme is in extreme demand as a ligand for hemoglobin. mtClpX may represent a new factor to consider in the etiology and treatment of disorders of heme biosynthesis.

Why might ALAS need a chaperone for PLP insertion? The free PLP cofactor is highly reactive and is maintained at a low concentration in the cell, near or below the dissociation constant for binding to many PLP-dependent enzymes (Cheung et al., 2003; Hamfelt, 1967). The very slow dissociation rate of PLP from its covalent attachment in active sites allows complex formation under these conditions, but spontaneous holoenzyme formation is likely to be inefficient. Therefore, active mechanisms for conjugation of PLP with apoenzymes have long been postulated. For example, pyridoxal kinase can interact with some PLP-dependent enzymes to shuttle newly generated PLP into the active sites of these enzymes in vitro (Cheung et al., 2003; Kim et al., 1988), but it is not known whether this is a general mechanism used in vivo. There is no known pyridoxal kinase activity in the mitochondrion, which suggests that a different mechanism (such as the one described here for ALAS) might be needed to facilitate PLP conjugation within this organelle. Determining whether mtClpX acts more broadly among mitochondrial PLP-dependent enzymes to facilitate PLP binding, or exclusively on ALAS, will be important to our understanding of the maturation and function of this large and important class of enzymes.

The Hsp70/90 chaperone system actively promotes ligand binding for several other classes of proteins, the best studied of which is the glucocorticoid receptor (Kirschke et al., 2014; Pratt et al., 2008). The Hsp70/90 enzymes modulate the structure of their substrates by a different mechanism than AAA+ unfoldases. Hsp70 and Hsp90 bind to partially unfolded intermediate structures, and the ATPase cycle does not appear coupled to an unfolding power stroke like that described for AAA+ unfoldases (Russell and Matouschek, 2014; Saibil, 2013). The requirement for a AAA+ unfoldase for apoALAS activation suggests a fundamentally different mechanism for facilitating ligand insertion, which may be dictated by the structural features of the substrate. In contrast to the structurally unstable and aggregation-prone unliganded glucocorticoid receptor (Kirschke et al., 2014), both apo- and holoALAS are well-structured dimers. This difference in the structure and folding of the substrate proteins could explain a requirement for directed unfolding by a AAA+ unfoldase rather than trapping of partially unfolded intermediates by Hsp70/90 to accelerate PLP binding.

How does mtClpX accelerate PLP binding to ALAS? Multiple residues within the folded active site of ALAS form important contacts with PLP, and studies of the prototypical α -family PLP enzyme, aspartate aminotransferase, revealed that PLP is lost upon enzyme unfolding (Astner et al., 2005; Gong et al., 1996; Wu et al., 2003). Because mitochondrial proteins are unfolded by the mitochondrial translocation machinery during

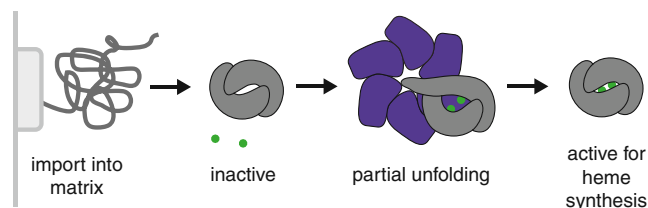


Figure 7. Model for mtClpX Activation of ALAS

ALAS is unfolded by the mitochondrial import machinery and refolds in the mitochondrial matrix. Newly folded ALAS binds PLP slowly on its own; partial unfolding by mtClpX renders the active site of ALAS more accessible to PLP, thus accelerating holoenzyme formation. Mitochondrial import machinery (light gray), ALAS (dark gray), mtClpX (purple), and PLP (green) are diagrammed.

import, newly imported and refolded ALAS is likely in the apo state. mtClpX is unlikely to extensively re-unfold ALAS as part of its activation mechanism, as this action would return the enzyme to a non-PLP binding form, although it is possible that the directionality or rate of mtClpX-mediated complete unfolding could promote refolding through a PLP-binding intermediate. One attractive model is that mtClpX locally unfolds or distorts the structure of ALAS to expose the buried active site and promote efficient PLP binding (Figure 7). The strong preference mtClpX exhibits for binding apoHem1 indicates that it must specifically recognize a structural feature or exposed motif that is specific to the apoenzyme to initiate this activity. Two distantly related AAA+ ATPases that function as dedicated activators for red- or green-type Rubisco stimulate a reverse event, a release of an inhibitory ligand (Mueller-Cajar et al., 2011; Wang and Portis, 1992). The better characterized red-type activase has been proposed to trigger ligand release by partial unfolding of Rubisco (Mueller-Cajar et al., 2011; Wang and Portis, 1992). Although stimulating ligand release and stimulating cofactor association are reverse biological processes, they may be mechanistically related. It will be interesting to compare the characteristics of the unfoldase-induced structural alterations in these distantly related systems.

A limited-unfolding model for ALAS activation by ClpX is consistent with the lack of ALAS degradation by mtClpXP. If unfolding by mtClpX is sufficiently limited, then the substrate protein would never be translocated far enough through the mtClpXP complex to reach the proteolytic active sites within the chamber of mtClpP. Because substrates of protease-coupled AAA+ unfoldases are often identified or validated by their degradation, such unfolding without coupled degradation may be a much more widespread activity than indicated by the repertoire of known substrates.

How does ClpX discriminate between activation and degradation? ClpX is best understood as part of the ClpXP proteolytic machine, and all previously identified substrates of ClpX are subject to degradation by ClpP. For two substrates, the tetrameric MuA transposase and the dimeric plasmid replication factor TrfA, unfolding rather than degradation is the biologically crucial event, although both substrates can be degraded by ClpXP (Koniczny and Helinski, 1997; Mhammedi-Alaoui et al., 1994). The function of ClpX in these cases is to extract one subunit from the complex by complete unfolding, altering the conformation

of the remaining protein(s); the subsequent degradation by ClpP of the unfolded subunit, therefore, is not deleterious. In the case of ALAS, where the dimer must remain intact to function, ClpX must instead activate the specific molecule upon which it exerts force. Therefore, proteolysis by ClpP must not occur; this ClpP-independent action of ClpX is what we observe.

Within the context of the cell, AAA+ unfoldases might be specified for nonproteolytic functions by sublocalization or interaction with other binding partners that is mutually exclusive with binding to their cognate protease. For example, yeast mtClpX associates tightly with the mitochondrial inner membrane (van Dyck et al., 1998); this association might sterically block mtClpP association in the eukaryotes that encode it as well as facilitate efficient activation of ALAS newly imported across the membrane.

Such complete biochemical uncoupling of substrate unfolding from proteolysis has not been previously observed for any protease-coupled AAA+ unfoldase, but two recently described activities of AAA+ proteases are informative for how this may be accomplished. The *C. crescentus* DNA polymerase clamp loader subunit (DnaX) is partially degraded by ClpXP to produce a functional isoform (Vass and Chien, 2013). To trigger N-end rule substrate delivery to ClpAP from the *E. coli* adaptor protein ClpS, the ClpA translocation pore engages ClpS itself without causing ClpS degradation (Rivera-Rivera et al., 2014; Román-Hernández et al., 2011). In these cases, local structure that is highly resistant to unfolding (in the case of DnaX and ClpS) and/or local peptide sequence that is poorly gripped by the unfoldase (DnaX) appear to release the protein from the grasp of the unfoldase, thereby attenuating or preventing degradation. Both of these types of elements also delimit the degraded region in two transcription factors that are processed by the 26S proteasome (Tian et al., 2005). Limited unfolding and translocation of ALAS by mtClpX, such as we propose, may be dictated by similar sequence and/or structural elements within ALAS. Defining this signal will help to delineate an emerging set of rules by which the fates of AAA+ unfoldase and protease substrates are determined.

EXPERIMENTAL PROCEDURES

Statistics

Error bars indicate SD, calculated from at least three biological replicates. The p values were calculated using Student's t test, unless otherwise indicated.

Yeast Strain Construction and Culture

Strains used in this study are listed in Table S1. Yeast genes were modified at chromosomal loci using standard homologous recombination techniques. Yeast strains were grown for experimental purposes in synthetic defined medium (YNB + CSM [yeast nitrogen base plus complete supplement mixture], Sunrise Sciences) with 2% dextrose at 30°C with shaking, unless otherwise indicated.

Heme and ALA Measurement

Cellular heme levels were monitored by porphyrin fluorescence and ^{55}Fe labeling. ALA levels in cell extracts were quantified using modified Ehrlich's reagent. ALA production by purified proteins was quantified using modified Ehrlich's reagent or by a NAD-coupled assay as indicated. See Extended Experimental Procedures for details.

Metabolic Profiling

Metabolite extracts were made by rapid vacuum filtration of yeast liquid culture, followed by incubation of the cell-laden filter in extraction solvent

(40% acetonitrile, 40% methanol, 20% water). Extracts were analyzed for relative metabolite levels by LC-MS. See [Extended Experimental Procedures](#) for further details.

Protein Purification and Biochemical Assays

mtClpX, mtClpP, and ALAS proteins were recombinantly produced in their mature forms (lacking the mitochondrial presequence). Biochemical assays for ALAS activity and PLP reconstitution were performed at 30°C in 25 mM HEPES (pH 7.6), 5 mM MgCl₂, and 10% glycerol with ATP regenerating system (5 mM creatine phosphate and 50 mg/ml creatine kinase, with 2 mM ATP when indicated), supplemented with 100 mM KCl (Hem1) or 130 mM KCl and 0.75 mg/ml BSA (ALAS2). 50 μM PLP was included in reconstitution experiments. Protein concentrations during PLP reconstitution were as follows: 3 μM apoHem1, 2 μM Mx1 (hexamer), 5 μM apoALAS2, 2 μM mouse ClpX (hexamer), 3 μM human ClpP-His₆ (14-mer). For detailed procedures, see [Extended Experimental Procedures](#).

Zebrafish Maintenance and Studies

Wild-type (AB), dino (*din^{tr250}*) ([Hammerschmidt et al., 1996](#)), and Tg(globin-LCR:eGFP) ([Ganis et al., 2012](#)) zebrafish (*Danio rerio*) were maintained, bred, and staged according to standard methods ([Lawrence et al., 2003](#)). Zebrafish studies were conducted with the approval of the Institutional Animal Care and Use Committee at Boston Children's Hospital. Embryos were stained for hemoglobinized cells with o-dianisidine as previously described ([Amigo et al., 2009](#)). Flow cytometry was used to quantify erythrocytes from embryos ([Cooney et al., 2013](#)).

Morpholino-Mediated Knockdown in Zebrafish

Splice site-blocking antisense morpholino oligomers ([Table S2](#)) were injected into one- and two-cell-stage embryos. Knockdown in morphant embryos was confirmed with real-time qRT-PCR using Taqman probes (Applied Biosystems). For details of ALA complementation, see [Extended Experimental Procedures](#).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, five figures, and two tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.04.017>.

ACKNOWLEDGMENTS

We thank S. Bell, J. Weissman, R. Vale, J. Nunnari, and N. Bradshaw for critical reading of the manuscript; L. Zon for the Tg(globin-LCR:eGFP) transgenic line; K. Hoffmeister for fluorescence-activated cell sorting (FACS) machine use; and C. Lawrence and his team for zebrafish husbandry. This work was supported by grants from the NIH (RO1 GM049224 to T.A.B., F32DK095726 to J.R.K., T32HL007574 and F32DK098866 to Y.Y.Y., and R01DK070838 and P01HL032262 to B.H.P.), the March of Dimes Foundation (6-FY09-289 to B.H.P.), and the Brazilian CAPES and FAPESP Foundations (to D.S.B.). K.Y.R. is the recipient of a Burroughs Wellcome Career Award in the Biomedical Sciences and of support from the William Randolph Hearst Foundation. T.A.B. is an employee of the Howard Hughes Medical Institute.

Received: September 4, 2014

Revised: January 15, 2015

Accepted: March 13, 2015

Published: May 7, 2015

REFERENCES

Al-Furokh, N., Kardon, J.R., Krüger, M., Szibor, M., Baker, T.A., and Braun, T. (2014). NOA1, a novel ClpXP substrate, takes an unexpected nuclear detour prior to mitochondrial import. *PLoS ONE* 9, e103141.

Amigo, J.D., Ackermann, G.E., Cope, J.J., Yu, M., Cooney, J.D., Ma, D., Langer, N.B., Shafizadeh, E., Shaw, G.C., Horsely, W., et al. (2009). The role

and regulation of friend of GATA-1 (FOG-1) during blood development in the zebrafish. *Blood* 114, 4654–4663.

Arrese, M., Carvajal, E., Robison, S., Sambunaris, A., Panek, A., and Mattoon, J. (1983). Cloning of the δ-aminolevulinic acid synthase structural gene in yeast. *Curr. Genet.* 7, 175–183.

Astner, I., Schulze, J.O., van den Heuvel, J., Jahn, D., Schubert, W.-D., and Heinz, D.W. (2005). Crystal structure of 5-aminolevulinic acid synthase, the first enzyme of heme biosynthesis, and its link to XLSA in humans. *EMBO J.* 24, 3166–3177.

Bogenhagen, D.F., Rousseau, D., and Burke, S. (2008). The layered structure of human mitochondrial DNA nucleoids. *J. Biol. Chem.* 283, 3665–3675.

Brownlie, A., Donovan, A., Pratt, S.J., Paw, B.H., Oates, A.C., Brugnara, C., Witkowska, H.E., Sassa, S., and Zon, L.I. (1998). Positional cloning of the zebrafish sauterne gene: a model for congenital sideroblastic anaemia. *Nat. Genet.* 20, 244–250.

Burton, R.E., Baker, T.A., and Sauer, R.T. (2003). Energy-dependent degradation: Linkage between ClpX-catalyzed nucleotide hydrolysis and protein-substrate processing. *Protein Sci.* 12, 893–902.

Camaschella, C. (2009). Hereditary sideroblastic anemias: pathophysiology, diagnosis, and treatment. *Semin. Hematol.* 46, 371–377.

Cheung, P.Y., Fong, C.C., Ng, K.T., Lam, W.C., Leung, Y.C., Tsang, C.W., Yang, M., and Wong, M.S. (2003). Interaction between pyridoxal kinase and pyridoxal-5-phosphate-dependent enzymes. *J. Biochem.* 134, 731–738.

Cooney, J.D., Hildick-Smith, G.J., Shafizadeh, E., McBride, P.F., Carroll, K.J., Anderson, H., Shaw, G.C., Tamplin, O.J., Branco, D.S., Dalton, A.J., et al. (2013). Teleost growth factor independence (gfi) genes differentially regulate successive waves of hematopoiesis. *Dev. Biol.* 373, 431–441.

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L., Toufighi, K., Mostafavi, S., et al. (2010). The genetic landscape of a cell. *Science* 327, 425–431.

Eliot, A.C., and Kirsch, J.F. (2004). Pyridoxal phosphate enzymes: mechanistic, structural, and evolutionary considerations. *Annu. Rev. Biochem.* 73, 383–415.

Ganis, J.J., Hsia, N., Trompouki, E., de Jong, J.L., DiBiase, A., Lambert, J.S., Jia, Z., Sabo, P.J., Weaver, M., Sandstrom, R., et al. (2012). Zebrafish globin switching occurs in two developmental stages and is controlled by the LCR. *Dev. Biol.* 366, 185–194.

Girvan, H.M., and Munro, A.W. (2013). Heme sensor proteins. *J. Biol. Chem.* 288, 13194–13203.

Gollub, E.G., Liu, K.P., Dayan, J., Adlersberg, M., and Sprinson, D.B. (1977). Yeast mutants deficient in heme biosynthesis and a heme mutant additionally blocked in cyclization of 2,3-oxidosqualene. *J. Biol. Chem.* 252, 2846–2854.

Gong, J., Kay, C.J., Barber, M.J., and Ferreira, G.C. (1996). Mutations at a glycine loop in aminolevulinic acid synthase affect pyridoxal phosphate cofactor binding and catalysis. *Biochemistry* 35, 14109–14117.

Gottesman, S. (2003). Proteolysis in bacterial regulatory circuits. *Annu. Rev. Cell Dev. Biol.* 19, 565–587.

Guernsey, D.L., Jiang, H., Campagna, D.R., Evans, S.C., Ferguson, M., Kellogg, M.D., Lachance, M., Matsuoka, M., Nightingale, M., Rideout, A., et al. (2009). Mutations in mitochondrial carrier family gene SLC25A38 cause non-syndromic autosomal recessive congenital sideroblastic anemia. *Nat. Genet.* 41, 651–653.

Hamfelt, A. (1967). Pyridoxal phosphate concentration and aminotransferase activity in human blood cells. *Clin. Chim. Acta* 16, 19–28.

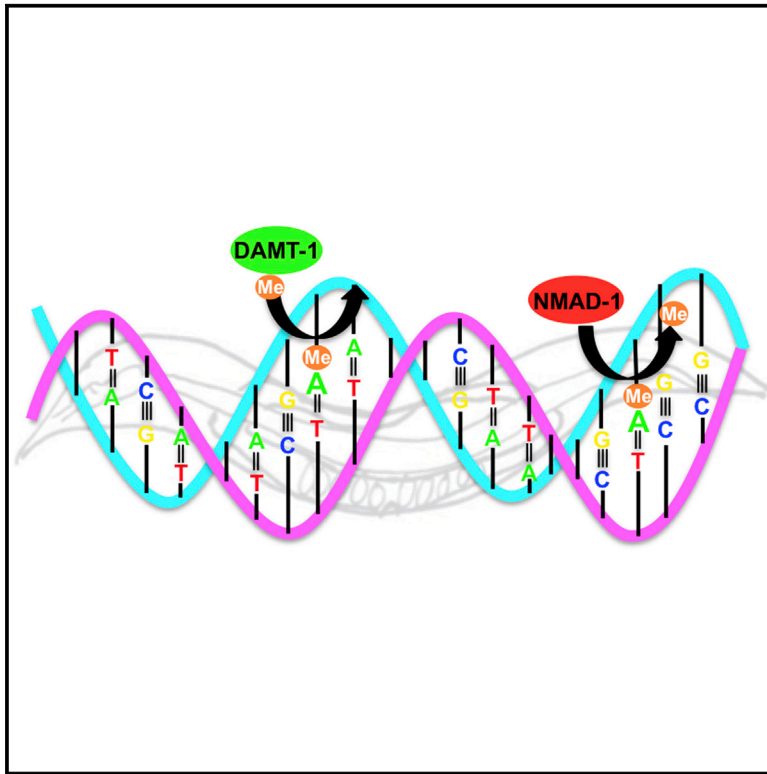
Hammerschmidt, M., Pelegri, F., Mullins, M.C., Kane, D.A., van Eeden, F.J., Granato, M., Brand, M., Furutani-Seiki, M., Haffter, P., Heisenberg, C.P., et al. (1996). dino and mercedes, two genes regulating dorsal development in the zebrafish embryo. *Development* 123, 95–102.

Hamza, I., and Dailey, H.A. (2012). One ring to rule them all: trafficking of heme and heme synthesis intermediates in the metazoans. *Biochim. Biophys. Acta* 1823, 1617–1632.

- Haynes, C.M., Yang, Y., Blais, S.P., Neubert, T.A., and Ron, D. (2010). The matrix peptide exporter HAF-1 signals a mitochondrial UPR by activating the transcription factor ZC376.7 in *C. elegans*. *Mol. Cell* 37, 529–540.
- Hoppins, S., Collins, S.R., Cassidy-Stone, A., Hummel, E., Devay, R.M., Lackner, L.L., Westermann, B., Schuldiner, M., Weissman, J.S., and Nunnari, J. (2011). A mitochondrial-focused genetic interaction map reveals a scaffold-like complex required for inner membrane organization in mitochondria. *J. Cell Biol.* 195, 323–340.
- Kang, S.G., Ortega, J., Singh, S.K., Wang, N., Huang, N.-N., Steven, A.C., and Maurizi, M.R. (2002). Functional proteolytic complexes of the human mitochondrial ATP-dependent protease, hClpX. *J. Biol. Chem.* 277, 21095–21102.
- Kasashima, K., Sumitani, M., and Endo, H. (2012). Maintenance of mitochondrial genome distribution by mitochondrial AAA+ protein ClpX. *Exp. Cell Res.* 318, 2335–2343.
- Kenniston, J.A., Baker, T.A., Fernandez, J.M., and Sauer, R.T. (2003). Linkage between ATP consumption and mechanical unfolding during the protein processing reactions of an AAA+ degradation machine. *Cell* 114, 511–520.
- Kim, Y.T., Kwok, F., and Churchich, J.E. (1988). Interactions of pyridoxal kinase and aspartate aminotransferase emission anisotropy and compartmentation studies. *J. Biol. Chem.* 263, 13712–13717.
- Kim, Y.I., Levchenko, I., Fraczowska, K., Woodruff, R.V., Sauer, R.T., and Baker, T.A. (2001). Molecular determinants of complex formation between Clp/Hsp100 ATPases and the ClpP peptidase. *Nat. Struct. Biol.* 8, 230–233.
- Kirschke, E., Goswami, D., Southworth, D., Griffin, P.R., and Agard, D.A. (2014). Glucocorticoid receptor function regulated by coordinated action of the Hsp90 and Hsp70 chaperone cycles. *Cell* 157, 1685–1697.
- Konieczny, I., and Helinski, D.R. (1997). The replication initiation protein of the broad-host-range plasmid RK2 is activated by the ClpX chaperone. *Proc. Natl. Acad. Sci. USA* 94, 14378–14382.
- Kořený, L., Sobotka, R., Kovářová, J., Gnipová, A., Flegontov, P., Horváth, A., Oborník, M., Ayala, F.J., and Lukeš, J. (2012). Aerobic kinetoplastid flagellate *Phytomonas* does not require heme for viability. *Proc. Natl. Acad. Sci. USA* 109, 3808–3813.
- Lawrence, J.M., Plank, L.R., and Lawrence, A.L. (2003). The effect of feeding frequency on consumption of food, absorption efficiency, and gonad production in the sea urchin *Lytechinus variegatus*. *Comp. Biochem. Physiol. A Mol. Integr. Physiol.* 134, 69–75.
- Lee, A.Y., St Onge, R.P., Proctor, M.J., Wallace, I.M., Nile, A.H., Spagnuolo, P.A., Jitkova, Y., Gronda, M., Wu, Y., Kim, M.K., et al. (2014). Mapping the cellular response to small molecules using chemogenomic fitness signatures. *Science* 344, 208–211.
- Martin, A., Baker, T.A., and Sauer, R.T. (2008a). Diverse pore loops of the AAA+ ClpX machine mediate unassisted and adaptor-dependent recognition of ssrA-tagged substrates. *Mol. Cell* 29, 441–450.
- Martin, A., Baker, T.A., and Sauer, R.T. (2008b). Pore loops of the AAA+ ClpX machine grip substrates to drive translocation and unfolding. *Nat. Struct. Mol. Biol.* 15, 1147–1151.
- Mhammedi-Alaoui, A., Pato, M., Gama, M.J., and Toussaint, A. (1994). A new component of bacteriophage Mu replicative transposition machinery: the *Escherichia coli* ClpX protein. *Mol. Microbiol.* 11, 1109–1116.
- Mueller-Cajar, O., Stotz, M., Wendler, P., Hartl, F.U., Bracher, A., and Hayer-Hartl, M. (2011). Structure and function of the AAA+ protein CbbX, a red-type Rubisco activase. *Nature* 479, 194–199.
- Novershtern, N., Subramanian, A., Lawton, L.N., Mak, R.H., Haining, W.N., McConkey, M.E., Habib, N., Yosef, N., Chang, C.Y., Shay, T., et al. (2011). Densely interconnected transcriptional circuits control cell states in human hematopoiesis. *Cell* 144, 296–309.
- Okano, S., Zhou, L., Kusaka, T., Shibata, K., Shimizu, K., Gao, X., Kikuchi, Y., Togashi, Y., Hosoya, T., Takahashi, S., et al. (2010). Indispensable function for embryogenesis, expression and regulation of the nonspecific form of the 5-aminolevulinic synthase gene in mouse. *Genes Cells* 15, 77–89.
- Pratt, W.B., Morishima, Y., and Osawa, Y. (2008). The Hsp90 chaperone machinery regulates signaling by modulating ligand binding clefts. *J. Biol. Chem.* 283, 22885–22889.
- Rivera-Rivera, I., Román-Hernández, G., Sauer, R.T., and Baker, T.A. (2014). Remodeling of a delivery complex allows ClpS-mediated degradation of N-degron substrates. *Proc. Natl. Acad. Sci. USA* 111, E3853–E3859.
- Román-Hernández, G., Hou, J.Y., Grant, R.A., Sauer, R.T., and Baker, T.A. (2011). The ClpS adaptor mediates staged delivery of N-end rule substrates to the AAA+ ClpAP protease. *Mol. Cell* 43, 217–228.
- Rottgers, K., Zufall, N., Guiard, B., and Voos, W. (2002). The ClpB homolog Hsp78 is required for the efficient degradation of proteins in the mitochondrial matrix. *J. Biol. Chem.* 277, 45829–45837.
- Russell, R., and Matouschek, A. (2014). Chance, destiny, and the inner workings of ClpXP. *Cell* 158, 479–480.
- Saibil, H. (2013). Chaperone machines for protein folding, unfolding and disaggregation. *Nat. Rev. Mol. Cell Biol.* 14, 630–642.
- Sassa, S. (2006). Modern diagnosis and management of the porphyrias. *Br. J. Haematol.* 135, 281–292.
- Sauer, R.T., Bolon, D.N., Burton, B.M., Burton, R.E., Flynn, J.M., Grant, R.A., Hersch, G.L., Joshi, S.A., Kenniston, J.A., Levchenko, I., et al. (2004). Sculpting the proteome with AAA(+) proteases and disassembly machines. *Cell* 119, 9–18.
- Schuldiner, M., Collins, S.R., Thompson, N.J., Denic, V., Bhamidipati, A., Punna, T., Ihmels, J., Andrews, B., Boone, C., Greenblatt, J.F., et al. (2005). Exploration of the function and organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* 123, 507–519.
- Shaw, G.C., Cope, J.J., Li, L., Corson, K., Hersey, C., Ackermann, G.E., Gwynn, B., Lambert, A.J., Wingert, R.A., Traver, D., et al. (2006). Mitoferrin is essential for erythroid iron assimilation. *Nature* 440, 96–100.
- Siddiqui, S.M., Sauer, R.T., and Baker, T.A. (2004). Role of the processing pore of the ClpX AAA+ ATPase in the recognition and engagement of specific protein substrates. *Genes Dev.* 18, 369–374.
- Tian, L., Holmgren, R.A., and Matouschek, A. (2005). A conserved processing mechanism regulates the activity of transcription factors Cubitus interruptus and NF-kappaB. *Nat. Struct. Mol. Biol.* 12, 1045–1053.
- Tian, Q., Li, T., Hou, W., Zheng, J., Schrum, L.W., and Bonkovsky, H.L. (2011). Lon peptidase 1 (LONP1)-dependent breakdown of mitochondrial 5-aminolevulinic acid synthase protein by heme in human liver cells. *J. Biol. Chem.* 286, 26424–26430.
- van Dyck, L., Dembowski, M., Neupert, W., and Langer, T. (1998). Mxc1p, a ClpX homologue in mitochondria of *Saccharomyces cerevisiae*. *FEBS Lett.* 438, 250–254.
- Vass, R.H., and Chien, P. (2013). Critical clamp loader processing by an essential AAA+ protease in *Caulobacter crescentus*. *Proc. Natl. Acad. Sci. USA* 110, 18138–18143.
- Volland, C., and Felix, F. (1984). Isolation and properties of 5-aminolevulinic synthase from the yeast *Saccharomyces cerevisiae*. *Eur. J. Biochem.* 142, 551–557.
- von der Haar, T. (2007). Optimized protein extraction for quantitative proteomics of yeasts. *PLoS ONE* 2, e1078.
- Wang, Z.Y., and Portis, A.R. (1992). Dissociation of ribulose-1,5-bisphosphate bound to ribulose-1,5-bisphosphate carboxylase/oxygenase and its enhancement by ribulose-1,5-bisphosphate carboxylase/oxygenase activase-mediated hydrolysis of ATP. *Plant Physiol.* 99, 1348–1353.
- Wintrobe, M.M., and Greer, J.P. (2004). Wintrobe's Clinical Hematology, Eleventh Edition (Philadelphia: Lippincott Williams & Wilkins).
- Wu, T.-H., Osés-Prieto, J.A., Iriarte, A., and Martínez-Carrion, M. (2003). Release of pyridoxal 5'-phosphate upon unfolding of mitochondrial aspartate aminotransferase. *Biochim. Biophys. Acta* 1647, 315–320.

DNA Methylation on N⁶-Adenine in *C. elegans*

Graphical Abstract



Authors

Eric Lieberman Greer,
Mario Andres Blanco, ..., Chuan He,
Yang Shi

Correspondence

eric.greer@childrens.harvard.edu
(E.L.G.),
yshi@hms.harvard.edu (Y.S.)

In Brief

Methylation is discovered to exist in *C. elegans* DNA on N⁶-adenines, along with a demethylase and putative methyltransferase. These enzymes are involved in trans-generational epigenetic signaling, raising the exciting possibility that this methyl mark may have an epigenetic role.

Highlights

- Identification of adenine N6 methylation (6mA) on DNA in *C. elegans*
- Examination of 6mA distribution
- Identification of a 6mA DNA demethylase and its role in epigenetic inheritance
- Identification of a potential 6mA DNA methylase and its role in epigenetic inheritance



DNA Methylation on N⁶-Adenine in *C. elegans*

Eric Lieberman Greer,^{1,2,5,*} Mario Andres Blanco,^{1,2,5} Lei Gu,^{1,2} Erdem Sendinc,^{1,2} Jianzhao Liu,³ David Aristizábal-Corrales,^{1,2} Chih-Hung Hsu,^{1,2} L. Aravind,⁴ Chuan He,³ and Yang Shi^{1,2,*}

¹Division of Newborn Medicine, Children's Hospital Boston, 300 Longwood Avenue, Boston, MA 02115, USA

²Department of Cell Biology, Harvard Medical School, Boston, MA 02115, USA

³Department of Chemistry and Institute for Biophysical Dynamics, Howard Hughes Medical Institute, The University of Chicago, Chicago, IL 60637, USA

⁴National Center for Biotechnology Information, National Library of Medicine, National Institutes of Health, Bethesda, MD 208943, USA

⁵Co-first author

*Correspondence: eric.greer@childrens.harvard.edu (E.L.G.), yshi@hms.harvard.edu (Y.S.)

<http://dx.doi.org/10.1016/j.cell.2015.04.005>

SUMMARY

In mammalian cells, DNA methylation on the fifth position of cytosine (5mC) plays an important role as an epigenetic mark. However, DNA methylation was considered to be absent in *C. elegans* because of the lack of detectable 5mC, as well as homologs of the cytosine DNA methyltransferases. Here, using multiple approaches, we demonstrate the presence of adenine N⁶-methylation (6mA) in *C. elegans* DNA. We further demonstrate that this modification increases trans-generationally in a paradigm of epigenetic inheritance. Importantly, we identify a DNA demethylase, NMAD-1, and a potential DNA methyltransferase, DAMT-1, which regulate 6mA levels and crosstalk between methylations of histone H3K4 and adenines and control the epigenetic inheritance of phenotypes associated with the loss of the H3K4me2 demethylase *spr-5*. Together, these data identify a DNA modification in *C. elegans* and raise the exciting possibility that 6mA may be a carrier of heritable epigenetic information in eukaryotes.

INTRODUCTION

An increasing number of complex phenotypes, including physical appearance (Cavalli and Paro, 1998; Morgan et al., 1999), energy metabolism (Benyshek et al., 2006), behavioral state (Dias and Ressler, 2014), and longevity (Greer et al., 2011; Rechavi et al., 2014), have been shown to be regulated in part by non-genetic information. The molecular nature of the epigenetic information that is transmitted from generation to generation is still incompletely understood. It has been postulated that anything in the zygote that is not the DNA sequence itself could carry this non-genetic information. This includes proteins, non-coding RNA, and modifications to both proteins and DNA in chromatin (Greer and Shi, 2012; Martin and Zhang, 2007; Moazed, 2011). Arguments have been made for each of these modes of epigenetic inheritance, and it is possible that a given mode of inheritance may play a larger role than others, depending on the paradigm of inheritance. One paradigm of epigenetic

inheritance in *C. elegans* involves mutation of the histone H3 lysine 4 dimethyl (H3K4me2) demethylase *spr-5* (Katz et al., 2009), which is an ortholog of the mammalian LSD1/KDM1A (Shi et al., 2004). The *spr-5* mutant worms initially do not exhibit phenotypes; however, after successive generations lacking this demethylase, they display a progressively increased infertility. This fertility decline is concomitant with a global increase in the activating histone mark H3K4me2 and decline in the repressive histone mark H3K9me3 (Greer et al., 2014; Katz et al., 2009; Kerr et al., 2014; Nottke et al., 2011). Despite the fact that early- and late-generation *spr-5* mutant worms should be genetically identical, late-generation *spr-5* mutant worms display altered phenotypes, most likely because of the inheritance of non-genetic information.

Previous studies searched for DNA modifications that carry epigenetic information in *C. elegans*. An early report performed high-performance liquid chromatography (HPLC) on *C. elegans* as they age and suggested that *C. elegans* have 5-methylcytosine (5mC) and that it accumulates with age (Klass et al., 1983). Other nematode species have also been reported to have 5mC (Gao et al., 2012); however, subsequent studies in *C. elegans* were unable to replicate this finding (Simpson et al., 1986). This lack of reproducibility, coupled with the fact that *C. elegans* do not contain homologs of the enzymes that add methyl moieties to cytosine—DNA (cytosine-5-)-methyltransferase 1 (DNMT1) or DNMT3—has led to the prevailing view that *C. elegans* do not possess DNA methylation (Wenzel et al., 2011). However, DNA is not only methylated at the fifth position of the pyrimidine ring of cytosines. Other DNA methylation events have been reported, including methylation of the exocyclic NH₂ groups at the sixth position of the purine ring in adenines (6mA) and at the fourth position of the pyrimidine ring in cytosines (4mC) (Iyer et al., 2011). In prokaryotes, 4mC and 6mA are primarily used for distinguishing self from foreign DNA (Iyer et al., 2011). These modifications are considered to be signaling or epigenetic modifications because they are predicted not to disrupt DNA base pairing (Iyer et al., 2011). Conversely, methylations of the first position of the purine ring in adenines (1mA) and the third position of the pyrimidine ring in cytosines (3mC) are considered DNA damage methylation events because they disrupt the hydrogen bonding with their base pairs. Additional DNA modifications have also been discovered or predicted in bacteria and eukaryotes (Iyer et al., 2011, 2013), but it remains

to be seen whether they are conserved across all species. Studies in eukaryotic organisms typically focus on 5mC and its role as an epigenetic modification (Koh and Rao, 2013; Martin and Zhang, 2007). However, it remains unknown whether DNA modifications such as 6mA and 4mC can also be used as epigenetic marks in eukaryotes and potentially even perpetuated through cell divisions and generations via the semi-conservative nature of DNA replication.

Here, we demonstrate that 6mA occurs in *C. elegans* DNA, is broadly distributed across the genome, and increases trans-generationally in *spr-5* mutant worms. We identify a 6mA DNA demethylase, NMAD-1, and show that deletion of *nmad-1* accelerates the progressive fertility defect phenotype of *spr-5* mutant worms. Conversely, deletion of *damt-1*, a potential 6mA DNA methyltransferase, reduces 6mA levels in worms and suppresses the progressive fertility defect of *spr-5* mutant worms. Additionally, we also identify reciprocal regulation between DNA 6mA and histone methylation. Our study identifies a new DNA modification in *C. elegans*, as well as regulators that control the dynamics of this modification, and advances 6mA as a potential carrier of non-genetic information across generations.

RESULTS

6mA Occurs in *C. elegans* and Increases Trans-generationally in *spr-5* Mutant Worms

To investigate whether any forms of DNA methylation are present in *C. elegans* and could be potential carriers of epigenetic memory in worms lacking *spr-5*, we extracted genomic DNA (gDNA) from whole worms and performed dot blot analysis on wild-type (WT) and late-generation *spr-5*(*by101*) mutant worms using a number of DNA modification-specific antibodies. Excitingly, we found that (1) 6mA, but not 5mC or 5hmC, was detectable in gDNA from WT worms and (2) the level of 6mA appears to be elevated in *spr-5* mutant worms (Figure S1A). To exclude the possibility that the detected 6mA is due to contamination from bacterial DNA, which contains 6mA, we used a bacterial food source that was deficient in the DNA adenine methyltransferase (Dam) and DNA cytosine methyltransferase (Dcm) enzymes, which are responsible for 6mA and 5mC modifications in bacteria, respectively (we confirmed that this mutant bacterial strain does not contain 6mA [Figure S1B]). To exclude the possibility that the detected 6mA was due to contaminating methylated RNA, we treated purified gDNA with enzymes targeting all major forms of RNA, including RNase A, RNase T1, and RNase H. We found that gDNA extracted from WT and late-generation *spr-5* mutant worms fed with *dam⁻dcm⁻* bacteria and treated with several RNases still exhibited detectable 6mA (Figure S1B). Furthermore, 6mA antibodies only detected very low signals from worm RNA dot blots, confirming that the observed 6mA DNA dot blot signals were not derived from potentially contaminating RNA in our genomic DNA preparations (Figures 1A and S1C). Lastly, we detected 6mA in worm gDNA samples using 6mA antibodies from two independent sources (Figures S1A and S1B).

We confirmed the specificity of the antibodies used in our dot blot analysis using a panel of unmethylated and premethylated DNA oligos (Figure S1D). Two 6mA antibodies (Synaptic

Systems and Megabase) recognized either single- or double-stranded 6mA- but not 3mC-containing oligos. The 6mA antibodies also recognized the non-denatured (double-stranded, ds), but not denatured (single-stranded, ss), 1mA (Figure S1D). Because the worm gDNA was denatured before being loaded onto blots, the 6mA antibody-detected signal was likely N⁶ adenine methylated DNA.

The elevation of 6mA in late generation *spr-5* mutant worms raises the possibility that 6mA might potentially play a role in transmitting heritable epigenetic information. Therefore, we next investigated whether the 6mA level changes in a trans-generational manner, as *spr-5* mutant worms have been shown to display a trans-generational increase in H3K4me2 level concomitant with trans-generational fertility defects (Greer et al., 2014; Katz et al., 2009). We found that 6mA increased in a trans-generational manner in *spr-5* mutant worms, regardless of worm culturing temperatures (Figure 1A). The magnitude of the increase in 6mA was variable across biological replicates, but the trend toward more 6mA in *spr-5* mutant worms was consistent.

To confirm that we were detecting 6mA, we turned to an antibody-independent approach, i.e., ultra-high-performance liquid chromatography coupled with a triple-quadrupole tandem mass spectrometry (UHPLC-MS/MS) analysis. We found that 6mA levels were variable from experiment to experiment in WT worms (occurring on between 0.01%–0.4% of adenines). However, 6mA levels were invariably elevated in the *spr-5*(*by101*) mutant worms, though the degree of upregulation differs from experiment to experiment (between 1.5- and 17-fold) and depends on the generation of worms assayed (Figure 1B and data not shown).

We initially noted that 1mA appeared to also increase in *spr-5* mutant worms as detected by the 1mA antibody (Figure S1C). However, the 1mA antibody recognizes both 1mA and 6mA oligos and therefore cannot distinguish the two modifications (Figure S1D), whereas UHPLC-MS/MS readily separates 1mA and 6mA (Figure S2). UHPLC-MS/MS analysis of WT and *spr-5* mutant worms gDNA typically failed to detect any 1mA in either strain (Figure S2B), indicating that the changes observed with our 1mA antibody likely reflected recognition of the elevated 6mA levels. On one occasion (out of more than ten trials) in which 1mA was detected by UHPLC-MS/MS, it was observed to be at similarly low levels in WT and *spr-5* mutant worms (Figure S1E).

We next investigated tissue distributions of 6mA by performing immunofluorescence (IF) on extracted germlines, embryos, and whole worms (Figures 1C, 1D, and S3A), which had been treated with RNases to remove potential RNA 6mA signal. We found 6mA present ubiquitously throughout the worm except for sperm nuclei (Figures 1C and S3A) and in every other cell in the worms' germline (Figure 1D). The absence of 6mA in sperm (Figure 1C) could reflect the high compaction of sperm chromatin (which might hamper the antibody accessibility) or could be indicative of a paternal erasure of 6mA. The IF signal likely represents 6mA, as pre-incubation of the antibodies with 6mA oligos, but not unmethylated oligos, abrogated the nuclear signal and resulted in a diffused, non-specific staining (Figure 1D). We also detected 6mA signal ubiquitously throughout the embryo (Figure 1C). Whereas 6mA was elevated in *spr-5* mutant worms

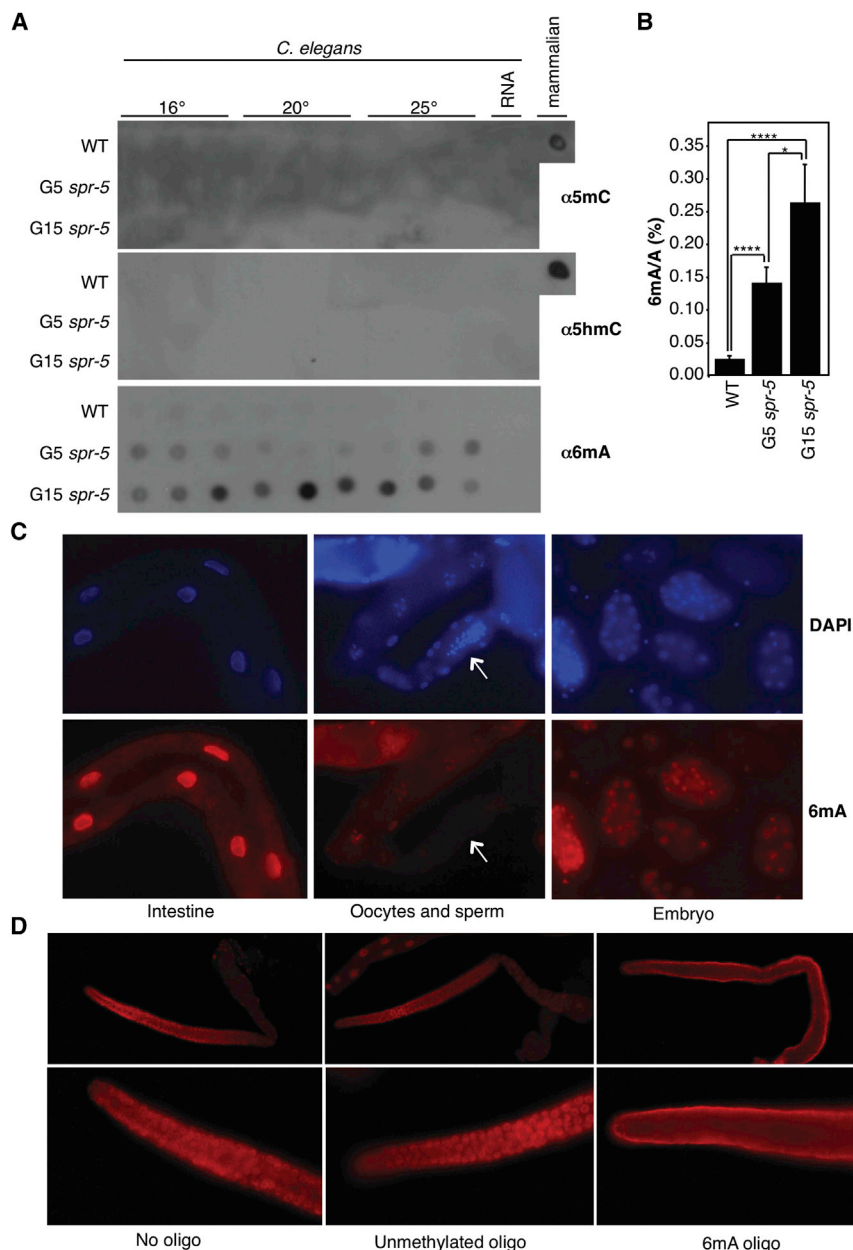


Figure 1. 6mA Occurs in *C. elegans* DNA and Increases across *spr-5* Generations

(A) Dot blots of three biological replicates of WT, generation 5, and generation 15 *spr-5(by101)* mutant worms grown at 16°, 20°, or 25° all show progressively elevated 6mA and lack detectable 5mC and 5hmC. 250 ng of gDNA are loaded per dot. Mammalian gDNA is used as a control for 5mC and 5hmC antibody strength.

(B) 6mA levels increase across generations of *spr-5(by101)* mutant worms as assessed by UHPLC-MS/MS. Each column represents the mean and SD of three to five biological replicates per group. **p* < 0.05 and *****p* < 0.0001.

(C) Immunofluorescence displays 6mA staining in the intestine, oocytes, and every cell of the embryo. Arrow indicates sperm.

(D) Immunofluorescence of wild-type extracted germlines shows 6mA in every nuclei. This staining was competed by a 6mA premethylated oligo but not by unmethylated oligos.

See also Figures S1, S2, and S3.

6mA Genomic Locations

For an initial investigation of 6mA genomic localization, we performed 6mA methylated DNA immunoprecipitation (Figure S5A), followed by sequencing (MeDIP-seq) on mixed-stage WT worms. MeDIP-seq identified 766 6mA peaks broadly distributed throughout the genome and evenly represented across major genomic features, except for a modest depletion in introns (Figure S5B). The most prevalent motif, AGAAGAAG AAGA, was present in 314 of the peaks identified (*p* = 1e−42, Figure S5C).

To more directly interrogate 6mA localization using an antibody-independent, base pair resolution approach, we carried out single-molecule real-time sequencing (SMRT sequencing), which not only identifies individual bases but also their modifications (Flusberg et al., 2010). We generated a SMRT sequencing dataset, using gDNA from mixed-stage, WT

(Figure S3A), 3mC and 1mA signals were essentially undetectable in germlines extracted from generation 20 (G20) *spr-5(by101)* mutant and WT worms (Figure S3B).

To determine whether 6mA might be associated with DNA damage, we performed dot blot analysis and stained gonads extracted from WT and DNA damage-deficient mutant strains. We found that deletion of the DNA damage genes, *xpa-1* (UV damage), *ercc-1* (nucleotide excision repair), and *sod-2* and *sod-3* (oxidative stress) did not lead to appreciably altered levels of 6mA (Figures S4A and S4B), nor did treatment with lethal doses of the DNA damaging agent methyl methanesulfonate (MMS) (Figure S4C). Together, these results suggest that 6mA is not a DNA damage-induced modification.

worms. To increase our read density, we merged our dataset with the publicly available *C. elegans* SMRT sequencing data generated by Pacific Biosciences (http://datasets.pacb.com.s3.amazonaws.com/2014/c_elegans/list.html). In this analysis, SMRT sequencing detected 6mA on 225,586 adenines—~0.7% of the total adenines in the worm genome—which is equivalent to 0.3% bulk adenine methylation, as some adenines were methylated 10% of the time, whereas others were methylated 90% of the time. This value (0.3%) is comparable to some of the UHPLC-MS/MS results (Figure 5E). SMRT sequencing does not distinguish 6mA versus 1mA, but 1mA is rarely above the level of detectability by UHPLC-MS/MS in worm gDNA. This suggests that the signals detected through SMRT sequencing

were 6mA (Figure 2A), although we cannot completely exclude the possibility that rare occurrences of 1mA could have been detected as 6mA in our SMRT sequencing analysis. Similar to the MeDIP-seq results, the SMRT sequencing analysis identified a broad distribution of 6mA across all chromosomes of the worm genome, with no one genomic feature being significantly enriched or depleted for 6mA (Figures 2B and 2C). Because lowly methylated regions usually include functional elements in mammalian cells (Stadler et al., 2011), we examined 6mA distribution (Figure 2C) by separating it into low (10%–20%, dark blue circle), middle (20%–80%, yellow circle), and high (80%–100%, red circle) categories and presenting the data in circo plot format in which concentric rings represent the density distributions of 6mA across the six worm chromosomes in the given category. We found that some lowly methylated regions appeared in dense clusters similar to lowly methylated 5mC (Figure 2C, innermost concentric circle). Notably, two sequence motifs were significantly associated with the presence of 6mA (Figure 2D): AGAA ($p = 1.9e-129$) and GAGG ($p = 5.1e-71$). Importantly, the AGAA motif identified by SMRT sequencing was also identified by MeDIP-seq (Figure S5C). Interestingly, the GAGG motif was most prevalent in sites that were frequently 6mA methylated (50%–100% methylation level), whereas the AGAA motif was most prevalent in infrequently 6mA methylated sites (10%–50% methylation level). The two 6mA motifs did not significantly differ in chromosomal distribution (Figure 2C, fourth concentric circle), though there were some regions that showed increased clustering density for each of the 6mA motifs (Figure 2C, outer rainfall plot). Notably, both motifs indicate that methylation at these sites occurs only on one of the strands, unlike the strong propensity for 5mC to occur in the context of CG doublets in various eukaryotes. Both SMRT sequencing and MeDIP-seq—which have been performed on mixed tissues and mixed-stage worms—confirmed the presence of 6mA in worm DNA across the genome and at similar sequence motifs.

Deletion of Potential Dealkylating Enzyme, *nmd-1*, Accelerates the Progressive Fertility Defect of *spr-5* Mutant Worms

To identify the enzymes responsible for the addition and removal of 6mA in *C. elegans*, we first examined the ALKB family of dealkylating enzymes, which have been shown in other species to remove methyl groups from DNA and RNA oxidatively, utilizing 2-oxoglutarate as a cofactor (Yi and He, 2013). Because 6mA levels increased across generations of *spr-5* mutant worms, we hypothesized that deletion of a 6mA demethylase would accelerate the trans-generational fertility defect of *spr-5* mutant worms. To determine whether any of the five *C. elegans* ALKB family members (Figure 3A) regulates 6mA, we first investigated whether knockdown or deletion of the family members had any effect on the progressive fertility defect of *spr-5* mutant worms. We found that knockdown of *Y51H7C.1*, *B0564.2*, *Y46G5A.35*, and *C14B1.10* had no effect on the fertility of WT or *spr-5* mutant worms (Figures 3B and S6A). Although we were unable to efficiently knock down the fifth ALKB family member, *F09F7.7* (Figure S6A), we obtained a worm strain carrying a deletion of *F09F7.7(ok3133)* and found that loss of *F09F7.7* accelerated the progressive fertility defect of *spr-5* mutant worms such that

the *spr-5;F09F7.7* double-mutant worms became completely sterile by generation 4 (Figure 3C). As a control, we found that, at a similar generation, the *spr-5* mutant worms did not display a significant fertility defect (Figure 3A; Greer et al., 2014; Katz et al., 2009). As a further control, we examined and found that *F09F7.7* mutants laid fewer eggs than WT worms (Figure 3C), but, importantly, this phenotype was not progressive (Figure 3D), suggesting that the acceleration of the progressive fertility defect of *spr-5* mutant worms is a result of a specific genetic interaction between *F09F7.7* and *spr-5*. These findings suggested that *F09F7.7* may act as a 6mA demethylase in vivo, which is further supported by the biochemical experiments discussed below. We thus renamed *F09F7.7* N6-methyl adenine demethylase 1 (*nmd-1*) to reflect this newly identified function.

NMAD-1 Demethylates 6mA In Vitro and In Vivo

To biochemically determine whether NMAD-1 was a 6mA demethylase, we glutathione S-transferase (GST) tagged and purified the protein and tested its demethylating activity in vitro. We found that two different isoforms of NMAD-1 were able to demethylate 6mA and 3mC oligos but not 1mA oligos in vitro (Figure 4A). To determine whether this demethylating activity was intrinsic to NMAD-1, we mutated the iron-chelating aspartic acid 186 in the catalytic domain of NMAD-1 to an alanine (D186A) and found that this mutation abrogated the ability of NMAD-1 to demethylate 6mA oligos (Figure 4B), suggesting that NMAD-1 possesses 6mA demethylase activity in vitro. We next investigated whether *nmd-1* mediates demethylation of both 6mA and 3mC in vivo. As shown in Figure 4C, *nmd-1* mutant worms showed elevated levels of 6mA, but not 3mC. This elevated 6mA was further confirmed by UHPLC-MS/MS (Figure 4D). Together, these results suggest that NMAD-1 is primarily a 6mA demethylase in vivo, although recombinant NMAD-1 protein can demethylate both 6mA and 3mC in vitro.

Deletion and Overexpression of the Potential Methyltransferase *damt-1* Decreases and Increases 6mA Levels In Vivo and in Tissue Culture, Respectively

We next sought to identify enzymes that mediate adenine N⁶-methylation in *C. elegans*. Although candidate 6mA DNA methyltransferases have been identified in chlorophyte algae, ciliates, some fungi, and certain other eukaryotic lineages (Iyer et al., 2011, 2014), none have been identified in Metazoa thus far. Although the eukaryotic candidate 6mA methyltransferases belong to multiple distinct methylase lineages (Iyer et al., 2011), the most widespread versions belong to the MTA-70 family exemplified by the yeast mRNA adenine methylase complex Ime4/Kar4 (Anantharaman et al., 2002; Clancy et al., 2002). These enzymes have evolved from m.MunI-like 6mA DNA methyltransferases of bacterial restriction-modification systems (Iyer et al., 2011) and are typified by a C-terminal circularly permuted methyltransferase domain fused to a distinctive N-terminal, predicted α -helical domain with a strongly positively charged segment. *C. elegans* has one representative of this family—the gene *C18A3.1*, which is conserved across eukaryotes, including humans, plants, basal fungi, certain amoebozoans, and stamenopiles and can be distinguished by phylogenetic analysis from Ime4 and Kar4 that are absent in *C. elegans* (Figures 5A and S6B). The orthologs of *C18A3.1*

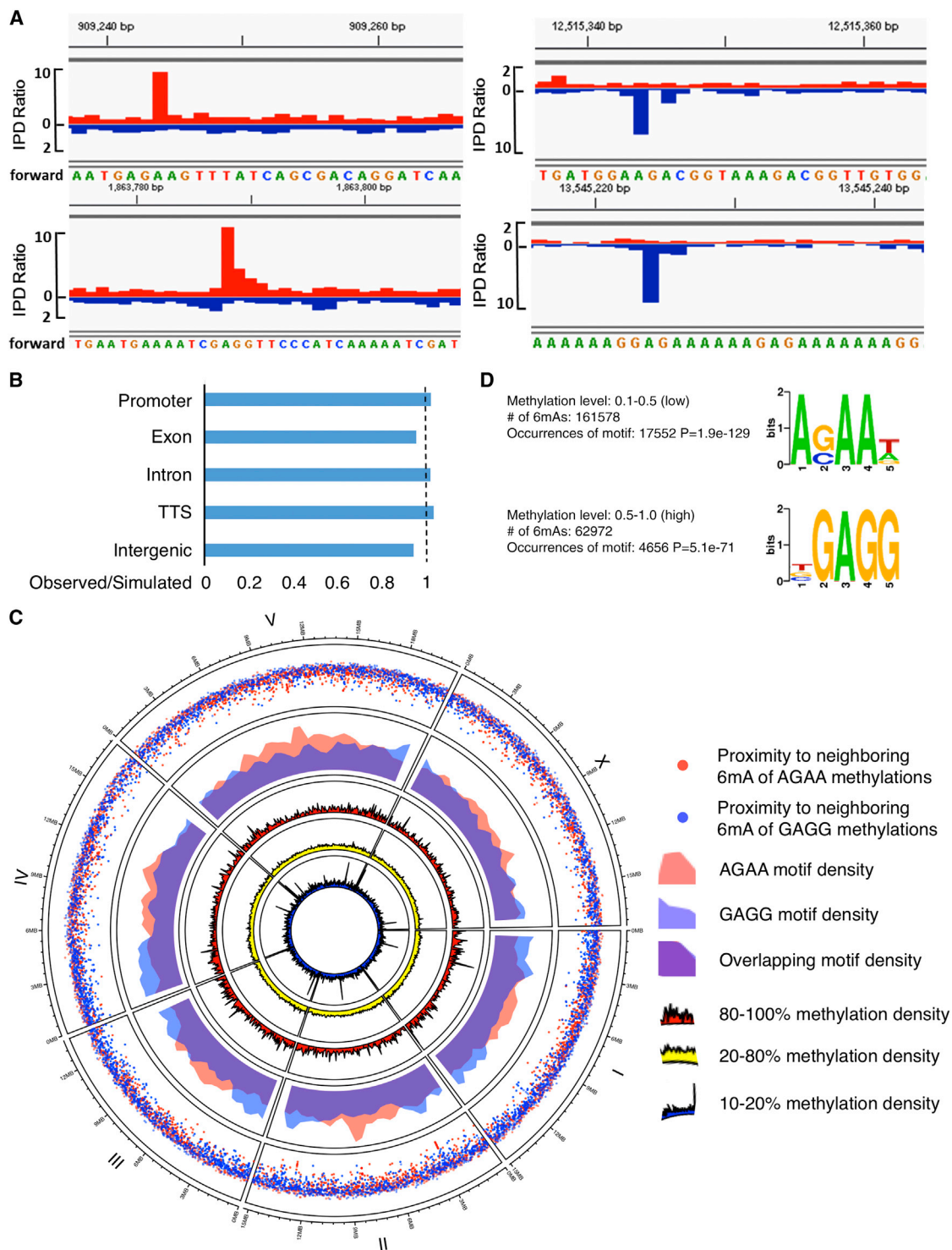


Figure 2. 6mA Genomic Location

(A) Representative interpulse duration (IPD) ratios of SMRT sequencing data of mixed-stage WT worms. IPD ratio is defined as the change in IPD distribution in the sample compared to unmodified bases. Red, positive strand; blue, negative strand.

(B) Comparison of observed versus simulated distributions of 6mA across the *C. elegans* genome indicates that 6mA is not enriched or depleted in any major genomic feature. A permutation was used to calculate the average of 10,000 simulations for comparison to the observed data.

(C) Circos plots of 6mA and motif distributions; three inner rings: 6mA density normalized to adenines in each bin of 6mAs within different methylation fractions. Red, yellow, and blue represent highly methylated (80%–100%), intermediate (20%–80%), and lowly methylated (10%–20%) 6mA, respectively. The middle ring

(legend continued on next page)

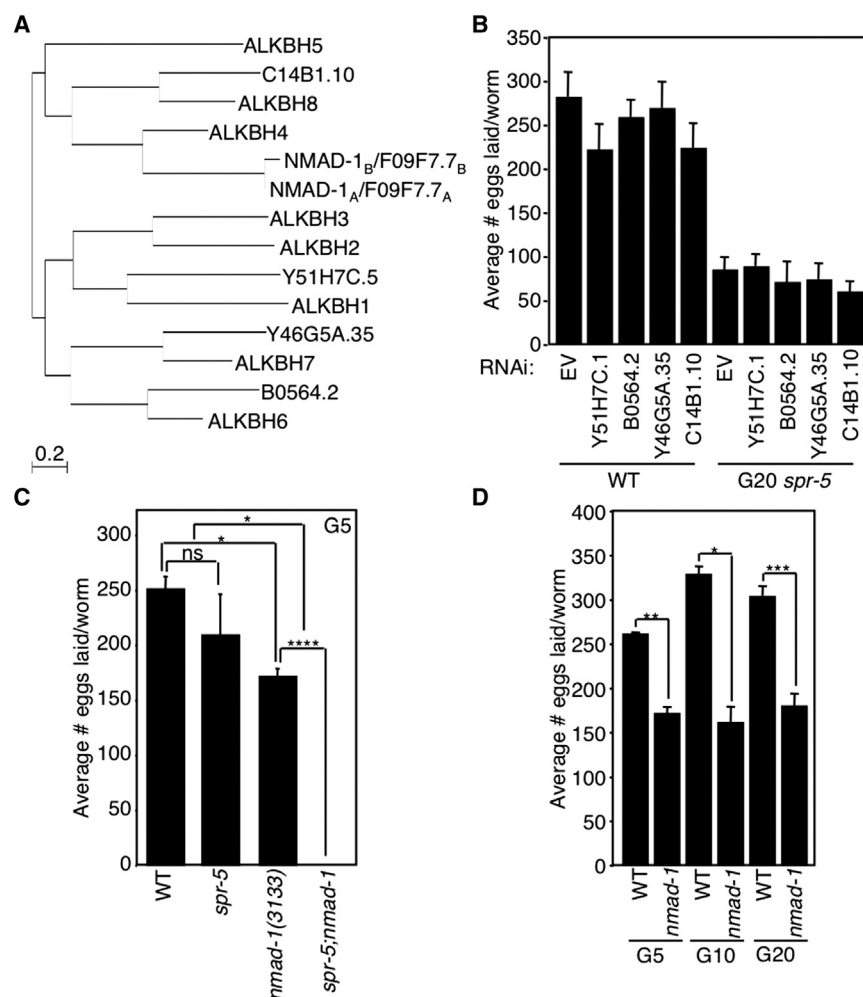


Figure 3. Deletion of *nmad-1* Accelerates the Progressive Fertility Defect of *spr-5* Mutant Worms

(A) Phylogeny tree of human and *C. elegans* ALKBH family members.

(B) Knockdown of 4 of the ALKBH family members has no effect on egg laying of WT and *spr-5* mutant worms treated for 20 generations with bacteria expressing the specific dsRNAs. Knockdown efficiency was tested by real-time RT-PCR (Figure S6A).

(C) Early-generation (G5) *spr-5* mutant worms do not display significant fertility defects, but when combined with *nmad-1* deletion, these worms become sterile by generation 4. Each bar represents the mean \pm SEM of three independent experiments.

(D) *nmad-1* mutants lay fewer eggs than WT worms but do not display a progressive fertility decline. Each bar represents the mean \pm SEM of two to six independent experiments. *p < 0.05, **p < 0.01, ***p < 0.001, and ****p < 0.0001; ns, not significant.

C18A3.1, we mutated amino acids in the N6A methyltransferase signature (DPPW) important for substrate recognition and catalytic activity (Iyer et al., 2011) and found that mutation of DPPW to APPA in the catalytic domain ablated the 6mA induction in SF9 gDNA (Figures 5C and S6C). This result suggests that C18A3.1 (renamed *damt-1* for DNA N6 adenine methyltransferase 1) is itself a 6mA methyltransferase, although we cannot rule out the less likely possibility that C18A3.1 expression in insect cells

form a distinct clade, separated from the mRNA methylase complex clade, within the primary eukaryotic radiation of the MTA-70 family. *C. elegans* also lack the transposon-encoded 6mA DNA methylase domains, which are found in related nematodes like *C. remanei*. These observations suggest that C18A3.1 is the primary 6mA DNA methylase candidate in *C. elegans*.

We investigated whether C18A3.1 could methylate the sixth position of adenines, but due to its high hydrophobicity, we were unable to purify this protein from bacterial or insect cells in sufficient quantities to study its activity in vitro. However, when we analyzed the gDNA isolated from SF9 cells expressing full-length C18A3.1 or the catalytic domain of C18A3.1 alone, we found that 6mA levels were elevated compared to DNA from insect cells that do not express C18A3.1 (Figure 5B). To determine whether this potential methylating activity was intrinsic to

coincidentally activated an endogenous insect cell enzyme that is responsible for the observed 6mA. To determine whether DAMT-1 was a 6mA methyltransferase in vivo, we knocked down *damt-1* in WT worms and found decreased 6mA but not 3mC levels in the extracted gDNA (Figures 5D and 5E). *damt-1* knockdown also decreased 6mA levels in *spr-5*(by101) mutant worms to similar levels as in WT worms (Figure 5F). Taken together, these data suggest that DAMT-1 is a 6mA methyltransferase in *C. elegans*.

Deletion of *damt-1* Suppresses the Trans-generational Phenotypes of *spr-5* Mutant Worms

If DAMT-1 is a 6mA methyltransferase, then we would expect that its knockdown or deletion would suppress the trans-generational phenotypes of *spr-5* mutant worms. Indeed, knockdown of

shows AGAA (red) and GAGG (blue) motif densities, with purple indicating the overlap. The outer ring (rainfall plot) shows the distribution of inter-distance between each two adjacent 6mAs in the same motif. Red dots represent 6mAs in AGAA motif, and blue dots represent 6mAs in GAGG motif; increasing vertical distance toward the center of the circle indicates increasing local density of 6mA occurrences.

(D) SMRT sequencing identified two motifs associated with 6mA. AGAA and GAGG are associated with low- and high-percentage 6mA, respectively. Methylation level refers to the percentage of times (1.0 = 100%) a given A in the sample population was read as methylated by SMRT sequencing.

See also Figure S5 for 6mA MeDIPseq.

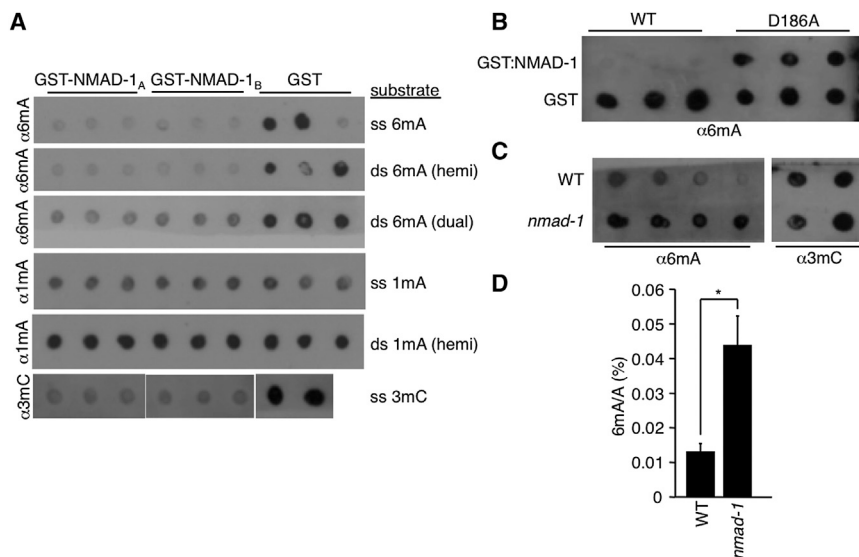


Figure 4. NMAD-1 Demethylates 6mA In Vitro and In Vivo

(A) Two different isoforms of NMAD-1 demethylate ss-denatured and ds-non-denatured (hemi or dual methylated) oligos premethylated at 6mA and 3mC but not 1mA.

(B) Mutation of the catalytic domain of NMAD-1 abrogates the ability of NMAD-1 to demethylate 6mA premethylated oligos.

(C) *nmad-1* mutants have elevated levels of 6mA without detectable changes in 3mC levels. Each dot represents 250 ng of DNA of independent biological replicates.

(D) *nmad-1* mutants have elevated levels of 6mA as assessed by UHPLC-MS/MS. Each bar represents the mean and SE of the mean of two independent biological replicates measured in duplicate. **p* < 0.05.

damt-1 for 20 generations partially suppressed the progressive fertility defect of *spr-5*(*by101*) mutant worms without affecting the fertility of WT worms (Figure 6A). Specifically, late-generation *spr-5* mutant worms on *damt-1* RNAi laid two to three times more eggs than late-generation *spr-5* mutant worms on bacteria containing an empty RNAi vector (Figure 6A). Similarly, a genetic deletion (*gk961032*) that removes the entirety of *damt-1* and a portion of the nearby Ras GTPase superfamily gene *rab-3* had no effect on egg laying by itself but suppressed the progressive fertility defect of *spr-5*(*by134*) mutant worms at generations 10, 17, 20, and 26 (Figure 6B and data not shown). *damt-1* knockdown also eliminated the fertility defect of the *nmad-1* mutant worms, suggesting that DAMT-1 functions to counteract the activity of the 6mA demethylase, NMAD-1, in vivo (Figure 6C). Collectively, these data suggest that DAMT-1 is a 6mA methyltransferase that suppresses the trans-generational phenotypes of *spr-5* mutant worms.

Crosstalk between H3K4me2 and 6mA

As discussed earlier, we initially observed an increase in 6mA levels in the histone H3K4me1/me2 demethylase mutant *spr-5*. Conversely, we found that deletion of the potential 6mA methyltransferase, *damt-1*, reduced the elevated H3K4me2 levels of *spr-5* mutant worms (Figures 7A, S7A, and S7B). Furthermore, we found that knockdown of the H3K9me binding protein *eap-1*, which reduces H3K4me2 levels in *spr-5* mutant worms (Greer et al., 2014), also reduced the levels of 6mA in *spr-5* mutant worms (Figures 7B and S7C). Collectively, these findings suggest reciprocal regulation of H3K4 and adenine N⁶ methylation and crosstalk between regulators that control adenine and histone methylation.

DISCUSSION

To date, 6mA has primarily been studied in prokaryotes, where it has been shown as a mark to discriminate invasive DNA (Arber and Dussoix, 1962; Meselson and Yuan, 1968). However, prokaryotic 6mA also functions as a binding platform and influences gene expression (Braun and Wright, 1986; Han et al., 2004). 6mA

has also been reported in more ancient eukaryotes such as ciliates, in which it is observed in the macro (somatic) and not in the micro (germline) nucleus, highlighting its potential function in a broad range of biological contexts (Gutiérrez et al., 2000). Both fungi and animals are known to undergo methylation of adenosine in mRNA, with 6mA influencing mRNA stability (Fu et al., 2014) and RNA splicing (Dominissini et al., 2012; Liu et al., 2015). However, whether 6mA is present in DNA of Metazoa has been unclear, and it has been widely assumed that 5mC, rather than 6mA, plays a primary role as the key carrier of epigenetic information on DNA in these organisms (Wion and Casadesús, 2006). Importantly, this study not only identifies the presence of 6mA in *C. elegans* but also raises the exciting possibility that this modification may play a role in carrying and transmitting epigenetic information across generations, and that, in addition to 5mC, 6mA may also be used across eukaryotes as a potential epigenetic modification.

Our conclusion that 6mA is present in the *C. elegans* genome is supported by multiple lines of evidence. First, 6mA was detected by two independently developed 6mA-specific antibodies (Figures 1A and S1A). Second, 6mA was detected on the DNA of most cells throughout the worm by immunofluorescence (Figures 1C, 1D, S3, and S4A). Third, the presence of 6mA was also identified by an antibody-independent means, i.e., UHPLC-MS/MS, which showed that *C. elegans* genome possesses 6mA (Figure 1B). Fourth, two independent sequencing methods—direct, antibody-independent DNA sequencing using SMRT sequencing and the antibody-dependent MeDIPseq—both detected 6mA on *C. elegans* DNA (Figures 2 and S5). Although both sequencing methods have caveats about distinguishing between 1mA and 6mA, the DNA samples subjected to sequencing had undetectable 1mA (as determined by UHPLC-MS/MS), suggesting that the majority of the methylation events detected by SMRT sequencing likely represent 6mA. Finally, we also identified potential enzymatic machineries that mediate addition and removal of 6mA (Figures 4 and 5). Importantly, manipulation of these enzymes in vivo not only affects 6mA

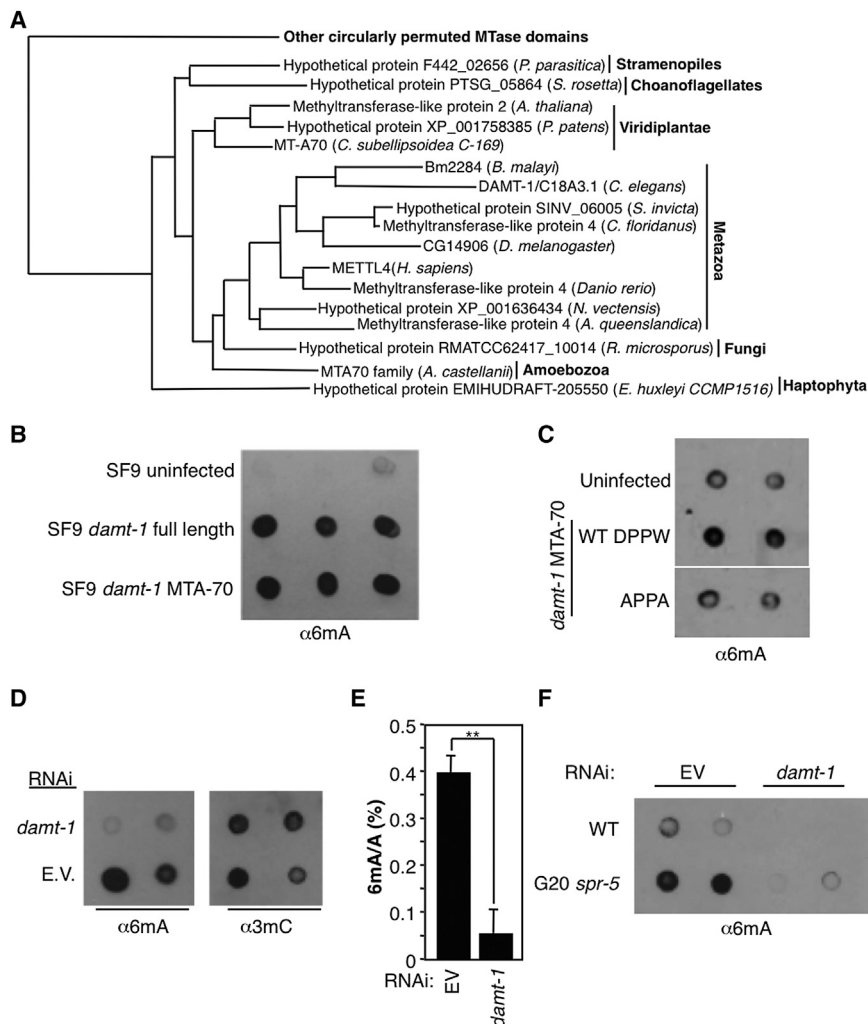


Figure 5. DAMT-1 Regulates 6mA Levels

(A) Phylogeny tree shows conservation of DAMT-1 in other eukaryotic species. Full tree and details of related clades are presented in Figure S6B.

(B) gDNA extracted from SF9 cells infected with full-length or the catalytic domain of *damt-1* show elevated levels of 6mA by dot blot.

(C) Mutation of the catalytic domain of DAMT-1 (DPPW to APPA) limits the increase in 6mA levels of infected SF9 cells. DAMT-1 expression is presented in Figure S6C.

(D) *damt-1* knockdown decreases 6mA without affecting detectable 3mC levels.

(E) *damt-1* mutants have decreased levels of 6mA as assessed by LC-MS/MS. Each bar represents the mean and SEM of three independent experiments of three biological replicates, each measured in duplicate. **p < 0.01.

(F) Generation 20 (G20) *spr-5* mutant worms show elevated 6mA levels compared with WT worms, and *damt-1* knockdown suppresses the elevated 6mA in *spr-5* mutant worms. Each dot represents 250 ng of gDNA of independent biological replicates.

levels but also impacts trans-generational epigenetic inheritance in *C. elegans* (Figures 3 and 6), raising the exciting and attractive possibility that 6mA may indeed carry epigenetic information.

Both SMRT sequencing and MeDIP-seq identified a broad 6mA genomic distribution with a common sequence motif but without a clear enrichment pattern; in contrast, 5mC distributions in mammals are highly tissue specific (Smith and Meissner, 2013). Given that worms of mixed developmental stages were used for sequencing, the possibility that 6mA may be enriched in specific genomic locations in a tissue-, cell-type-, or developmental-stage-specific manner remains, and such enrichment patterns may only emerge when DNA samples from specific cell types or developmental stages are analyzed.

Although DNA methylation may be a more efficient carrier of epigenetic information, it remains to be seen whether 6mA, H3K4me2, or some as-of-yet-unidentified mark carry the epigenetic information on their own or collaborate to transmit epigenetic information across generations in *C. elegans*. A recent study provided evidence that both the histone modification mark (H3K27me3) and the PRC2 machinery are transmitted across generations epigenetically (Gaydos et al.,

2014), implicating chromatin modifications as possible carriers of heritable non-genetic information. Interestingly, our study identified robust genetic interactions between the H3K4me1/2-specific demethylase SPR-5 and machineries that regulate 6mA—i.e., NMAD-1 and DAMT-1—in the regulation of trans-generational epigenetic inheritance. These results suggest crosstalk between 6mA and histone methylation and possible collaboration of these modifications in transmitting epigenetic information. Further evidence for this crosstalk was provided by the finding that knockdown of the H3K9me binding protein, *eap-1*, which reduces H3K4me2 levels in *spr-5* mutant worms (Greer et al., 2014), also decreases 6mA levels in *spr-5* mutant worms (Figure 7B). Conversely, deletion of the potential 6mA methyltransferase, *damt-1*, decreases H3K4me2 levels in *spr-5* mutant worms (Figure 7A). Consistent with the possibility of crosstalk between H3K4 and adenine N⁶ methylation regulation, analysis of the domain architectures of DNA N6A methyltransferases in eukaryotes, such as chlorophytes and fungi, showed that the DNA-modifying catalytic domain is fused to histone-recognition domains (Iyer et al., 2011, 2014).

At the present time, the molecular function of 6mA is still unclear. DNA methylation systems such as 6mA and 5mC are proposed to serve various functions, including protection of host genomes (Arber and Dussoix, 1962; Meselson and Yuan, 1968), silencing of transposable elements (Kato et al., 2003; Zemach and Zilberman, 2010), transcriptional silencing (Csankovszki et al., 2001; Sado et al., 2000; Stein et al., 1982), prevention of cryptic transcription in intragenic regions (Zemach et al., 2010), and heterochromatin state transitions (Saksouk et al.,

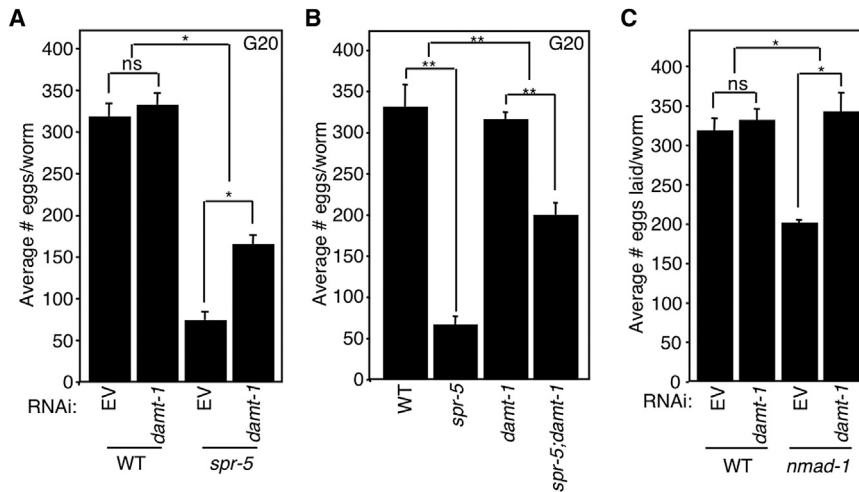


Figure 6. Deletion of *damt-1* Suppresses the Trans-generational Phenotypes of *spr-5* Mutant Worms

(A) *damt-1* knockdown has no effect on WT egg laying but partially suppresses the progressive fertility defect of *spr-5*(*by101*) mutant worms. (B) *damt-1* deletion has no effect on WT egg laying but partially suppresses the progressive fertility defect of *spr-5*(*by134*) mutant worms. (C) *damt-1* knockdown reverts the egg-laying defect of *nmad-1* mutant worms. All assays were performed at generation 20. Each bar represents the mean \pm SEM of three independent experiments. * $p < 0.05$ and ** $p < 0.01$; ns, not significant.

2014). A study conducted in *Chlamydomonas reinhardtii* (Fu et al., 2015 [this issue of Cell]) shows a correlation of 6mA modification with active gene transcription, suggesting a possible role in gene expression regulation. We observed that the absolute 6mA levels were variable from experiment to experiment and found that some environmental manipulations altered 6mA levels (data not shown). This raises the possibility that this modification could integrate environmental stimuli to regulate biological processes. Future studies will be required to fully explore the molecular function of 6mA in worms.

Finally, it will be informative to place 6mA regulation within a cellular pathway(s). In *Arabidopsis*, for example, the RNAi pathway feeds into 5mC regulation and heterochromatin formation and propagation (Law and Jacobsen, 2010; Teixeira et al., 2009; Wassenegger et al., 1994). Whether molecular pathways governing the trans-generational epigenetic inheritance of fertility and other phenotypes feed into 6mA regulation in *C. elegans* remains to be determined. It will be of significant interest to understand whether 6mA contributes to regulating the epigenome landscape that governs trans-generational epigenetic inheritance. Furthermore, given that orthologs of *damt-1* are widely conserved across eukaryotes, including mammals and other vertebrates, it will now be of great interest to investigate

which other eukaryotic species might also have 6mA in their DNA and in which biological contexts this modification is regulated and plays a biological function.

If other eukaryotes are found to have 6mA, it raises the exciting possibility that 6mA could carry epigenetic information in multiple paradigms of epigenetic inheritance.

EXPERIMENTAL PROCEDURES

Worm Strains

The N2 Bristol strain was used as the WT background. The following mutations were used in this study: LG1: *spr-5*(*by101*), *spr-5*(*by134*), *ercc-1*(*tm1981*), *xpa-1*(*mn157*), *sod-2*(*gk257*); LGII: *damt-1*(*gk961032*); LGIII: *nmad-1*(*ok3133*), LGX: *sod-3*(*tm760*). In this paper, mutant worms were backcrossed: *damt-1*, 5–7 times; *nmad-1*, 5–9 times. Worms were grown on *dam⁻ dcm⁻* bacteria (NEB C2925) in all experiments except for Figure S1A, where they were grown on OP50-1 bacteria.

Fertility Assays

From day 3 to day 8 post-hatching, 10 worms were placed on NGM plates with bacteria in triplicate (30 worms total per condition). Worms were grown at 20°C. After 24 hr, the adult worms were removed from each plate and placed on new plates. The numbers of eggs and hatched worms on the plate were counted. Statistical analyses of fertility were performed using two-way ANOVA tests with Bonferroni post-tests or t tests using mean and standard error values.

Worm gDNA Extraction

Worms were washed two times with M9 buffer. 250 μ l of worm genomic DNA lysis buffer (200 mM NaCl, 100 mM Tris-HCl [pH 8.5], 50 mM EDTA [pH 8.0],

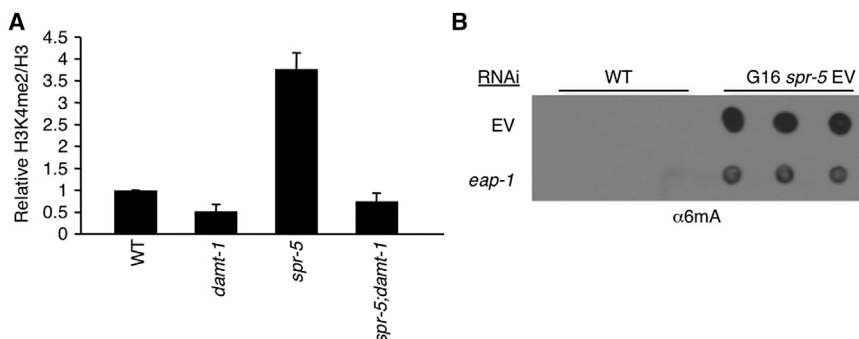


Figure 7. DNA Methylation and Histone Methylation Crosstalk

(A) Deletion of *damt-1* suppresses the elevated H3K4me2 levels of late-generation *spr-5*(*by134*) mutant worms. Each bar represents the mean \pm SEM of three independent experiments performed in biological duplicate. Image J was used to analyze the relative intensity of H3K4me2 compared to histone H3. Western blots corresponding to two of these experiments are shown in Figures S7A and S7B.

(B) Knockdown of H3K9me binding protein, *eap-1*, suppresses the elevated 6mA level detected in *spr-5* mutant worms as assessed by dot blots. A longer exposure showing 6mA levels in WT worms is shown in Figure S7C.

0.5% SDS) + proteinase K (0.1 mg/ml) was added. Worms were incubated at 65°C for 1 hr with occasional vortexing and then incubated at 95°C for 20 min. RNase A was added (0.1 mg/ml) and incubated at 37°C for 1 hr. 250 μ l of phenol:chloroform:isoamyl alcohol was added. Samples were mixed and then spun at 13,000 rpm at room temperature for 15 min. The aqueous phase was removed to a new tube, and phenol:chloroform:isoamyl alcohol extraction was repeated. To the aqueous phase, 25 μ l of 3M sodium acetate and 750 μ l of 100% EtOH were added and samples were placed at –80°C for at least 1 hr. Samples were spun at 13,000 rpm at 4°C for 30 min. The supernatant was removed. 350 μ l of cold 75% EtOH was added, and samples were again spun at 13,000 rpm for 10 min. The supernatant was discarded, and pellet was allowed to dry before being resuspended in TE (10 mM Tris-HCl, 1 mM EDTA [pH 8.0, final pH 7.5]). For samples presented in [Figures 7B, S1B, S4C, and S7C](#), purified gDNA was then treated with RNase A/T1 mix (Thermo Scientific) at a 1:20 dilution and RNaseH (NEB) at a 1:50 dilution for 1 hr at 37°C for 1 hr prior to subsequent re-purification starting with proteinase K digestion.

Dot Blot

Samples were diluted to 100 ng/ μ l and heated at 95°C for 10 min to denature DNA. Samples were immediately placed on ice for 5 min, and 250 ng were loaded per dot on Hybond + membranes. Membranes were allowed to air dry and placed in boxes with damp paper towels. DNA was then autocross-linked in a UV stratalinker 2400 (Stratagene) two times. The membrane was allowed to dry and then blocked for 1 hr in 5% milk TBS. Membranes were probed for 1 hr at room temperature or overnight at 4°C with primary antibody in 5% milk TBS. Blots were washed three times for 10 min with TTBS and then probed with secondary antibody in 5% milk for 1 hr at room temperature. Blots were washed three times for 10 min with TTBS, and ECL was applied and film was developed.

SMRT Sequencing

The raw data are from two parts: (1) our own data, uploaded into GEO (accession number GSE66504) and (2) from PacBio public database (http://datasets.pacb.com.s3.amazonaws.com/2014/c_elegans/list.html). Each of the raw data in bax.h5 format were first aligned to ce10 genome using pbalin in base modification identification mode. The polymerase kinetics information was further loaded after alignment by loadChemistry.py and loadPulses scripts. Then two post-aligned datasets were merged and sorted by using cmph5tools. Finally, the 6mA was identified using ipdSummary.py script. We then further filtered 6mAs with less than 50 \times coverage. For motif identification, we first separated the whole 6mAs into 10 groups based on their methylation level (methylation level ranges: 0%–10%; 10%–20%...90%–100%). For each 6mA, we then extracted 2bp from the upstream and downstream sequences. MEME-ChIP ([Machanic and Bailey, 2011](#)) was then used to identify motifs in each group. The genome-wide 6mA and motif profiles are generated from circlize ([Gu et al., 2014](#)). Part of the analysis was done by customized scripts in R, Python, and Perl.

Antibodies

The following antibodies were used: α 6mA (Synaptic Systems, 202 003), α 6mA (Megabase Research), α 5mC (Active Motif, 39649), α 5hmC (Active Motif 39769), α 3mC (Active Motif, 61111 and 61179), and α 1mA (Active Motif, custom). α 6mA (Megabase Research) was only used in [Figure S1A](#).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures and seven figures and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.04.005>.

AUTHOR CONTRIBUTIONS

E.L.G., M.A.B., and Y.S. conceived and planned the study and wrote the paper. E.L.G. produced [Figures 1A, 3, 5A, 5D, 5F, 6, 7A, S1C, S3A, S4B, S4C, S6A, S7A, and S7B](#). M.A.B. produced [Figures 1A, 4A, 4B, 4C, S1A, S1B, S1D, and S5A](#). L.G. performed bioinformatics analysis presented in [Figures 2,](#)

[S5B, and S5C](#). E.S. produced [Figures 5B, 5C, 7B, S6C, and S7C](#). J.L. performed UHPLC-MS/MS experiments shown in [Figures 1B, 4D, 5E, S1E, and S2](#) and was advised by C.H. D.A.-C. produced [Figures 1C, 1D, S3B, and S4A](#). C.-H.H. performed protein purifications and DNA methylation assays. L.A. identified *damt-1* bioinformatically and produced [Figures 5A and S6B](#).

ACKNOWLEDGMENTS

We thank members of the Shi lab for helpful discussions, Elizabeth Pollina for feedback on the manuscript, and Madeline Schuck and LaVondea Elow for technical and administrative support. We thank the Caenorhabditis Genetics Center, which is funded by NIH Office of Research Infrastructure Programs (P40OD010440) for *C. elegans* strains, and the Tufts University Core Facility Genomics Core and the University of Massachusetts Medical School Deep Sequencing Core for MeDIP-seq and SMRT sequencing, respectively. E.L.G. was supported by a Helen Hay Whitney postdoctoral fellowship and a National Institute on Aging of the NIH grant (K99AG043550). M.A.B. was supported by an NIH NRSA postdoctoral fellowship (1F32CA180450-01) and is currently supported by a Special Fellow award from the Leukemia & Lymphoma Society (3353-15). L.A. was supported by funds of the Intramural Research Program of the National Library of Health, NIH, and U.S. department of Health and Human Services. C.H. is a Howard Hughes Medical Institute investigator. This work was supported by NIH grants to Y.S. (GM058012, CA118487, and MH096066) and E.L.G. (K99AG043550), by an Ellison Foundation Senior Scholar Award to Y.S., and by a Samuel Waxman Cancer Research Foundation grant to Y.S. (SWCRF-1856). Y.S. is an American Cancer Society Research Professor. Y.S. is also a cofounder of Constellation Pharmaceuticals Inc. and is a member of its scientific advisory board.

Received: November 8, 2014

Revised: March 2, 2015

Accepted: March 31, 2015

Published: April 30, 2015

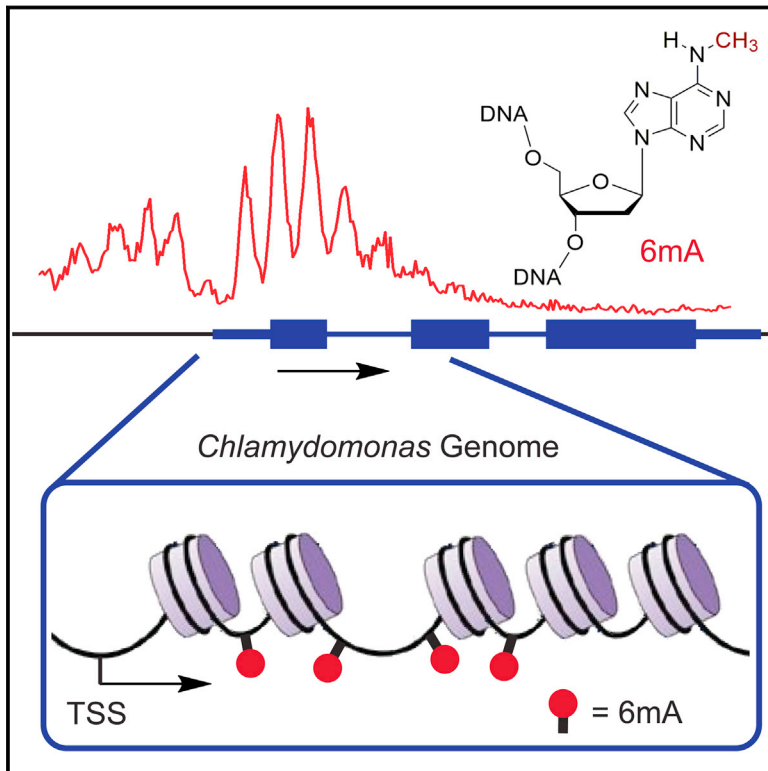
REFERENCES

- Anantharaman, V., Koonin, E.V., and Aravind, L. (2002). Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Res.* 30, 1427–1464.
- Arber, W., and Dussoix, D. (1962). Host specificity of DNA produced by *Escherichia coli*. I. Host controlled modification of bacteriophage lambda. *J. Mol. Biol.* 5, 18–36.
- Benyshek, D.C., Johnston, C.S., and Martin, J.F. (2006). Glucose metabolism is altered in the adequately-nourished grand-offspring (F3 generation) of rats malnourished during gestation and perinatal life. *Diabetologia* 49, 1117–1119.
- Braun, R.E., and Wright, A. (1986). DNA methylation differentially enhances the expression of one of the two *E. coli* dnaA promoters in vivo and in vitro. *Mol. Gen. Genet.* 202, 246–250.
- Cavalli, G., and Paro, R. (1998). The *Drosophila* Fab-7 chromosomal element conveys epigenetic inheritance during mitosis and meiosis. *Cell* 93, 505–518.
- Clancy, M.J., Shambaugh, M.E., Timpte, C.S., and Bokar, J.A. (2002). Induction of sporulation in *Saccharomyces cerevisiae* leads to the formation of N6-methyladenosine in mRNA: a potential mechanism for the activity of the IME4 gene. *Nucleic Acids Res.* 30, 4509–4518.
- Csankovszki, G., Nagy, A., and Jaenisch, R. (2001). Synergism of Xist RNA, DNA methylation, and histone hypoacetylation in maintaining X chromosome inactivation. *J. Cell Biol.* 153, 773–784.
- Dias, B.G., and Ressler, K.J. (2014). Parental olfactory experience influences behavior and neural structure in subsequent generations. *Nat. Neurosci.* 17, 89–96.
- Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., et al. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* 485, 201–206.

- Flusberg, B.A., Webster, D.R., Lee, J.H., Travers, K.J., Olivares, E.C., Clark, T.A., Korlach, J., and Turner, S.W. (2010). Direct detection of DNA methylation during single-molecule, real-time sequencing. *Nat. Methods* 7, 461–465.
- Fu, Y., Dominissini, D., Rechavi, G., and He, C. (2014). Gene expression regulation mediated through reversible m⁶A RNA methylation. *Nat. Rev. Genet.* 15, 293–306.
- Fu, Y., Luo, G.-Z., Chen, K., Deng, X., Yu, M., Han, D., Hao, Z., Liu, J., Lu, X., Dore, L.C., et al. (2015). N⁶-methyldeoxyadenosine marks active transcription start sites in *Chlamydomonas*. *Cell* 161, this issue, 879–892.
- Gao, F., Liu, X., Wu, X.P., Wang, X.L., Gong, D., Lu, H., Xia, Y., Song, Y., Wang, J., Du, J., et al. (2012). Differential DNA methylation in discrete developmental stages of the parasitic nematode *Trichinella spiralis*. *Genome Biol.* 13, R100.
- Gaydos, L.J., Wang, W., and Strome, S. (2014). Gene repression. H3K27me and PRC2 transmit a memory of repression across generations and during development. *Science* 345, 1515–1518.
- Greer, E.L., and Shi, Y. (2012). Histone methylation: a dynamic mark in health, disease and inheritance. *Nat. Rev. Genet.* 13, 343–357.
- Greer, E.L., Maures, T.J., Ucar, D., Hauswirth, A.G., Mancini, E., Lim, J.P., Benayoun, B.A., Shi, Y., and Brunet, A. (2011). Transgenerational epigenetic inheritance of longevity in *Caenorhabditis elegans*. *Nature* 479, 365–371.
- Greer, E.L., Beese-Sims, S.E., Brookes, E., Spadafora, R., Zhu, Y., Rothbart, S.B., Aristizabal-Corrales, D., Chen, S., Badeaux, A.I., Jin, Q., et al. (2014). A histone methylation network regulates transgenerational epigenetic memory in *C. elegans*. *Cell Rep.* 7, 113–126.
- Gu, Z., Gu, L., Eils, R., Schlesner, M., and Brors, B. (2014). circlize Implements and enhances circular visualization in R. *Bioinformatics* 30, 2811–2812.
- Gutiérrez, J.C., Callejas, S., Borniquel, S., and Martín-González, A. (2000). DNA methylation in ciliates: implications in differentiation processes. *Int. Microbiol.* 3, 139–146.
- Han, J.S., Kang, S., Kim, S.H., Ko, M.J., and Hwang, D.S. (2004). Binding of SeqA protein to hemi-methylated GATC sequences enhances their interaction and aggregation properties. *J. Biol. Chem.* 279, 30236–30243.
- Iyer, L.M., Abhiman, S., and Aravind, L. (2011). Natural history of eukaryotic DNA methylation systems. *Prog. Mol. Biol. Transl. Sci.* 101, 25–104.
- Iyer, L.M., Zhang, D., Burroughs, A.M., and Aravind, L. (2013). Computational identification of novel biochemical systems involved in oxidation, glycosylation and other complex modifications of bases in DNA. *Nucleic Acids Res.* 41, 7635–7655.
- Iyer, L.M., Zhang, D., de Souza, R.F., Pukkila, P.J., Rao, A., and Aravind, L. (2014). Lineage-specific expansions of TET/JBP genes and a new class of DNA transposons shape fungal genomic and epigenetic landscapes. *Proc. Natl. Acad. Sci. USA* 111, 1676–1683.
- Kato, M., Miura, A., Bender, J., Jacobsen, S.E., and Kakutani, T. (2003). Role of CG and non-CG methylation in immobilization of transposons in *Arabidopsis*. *Curr. Bio.* 13, 421–426.
- Katz, D.J., Edwards, T.M., Reinke, V., and Kelly, W.G. (2009). A *C. elegans* LSD1 demethylase contributes to germline immortality by reprogramming epigenetic memory. *Cell* 137, 308–320.
- Kerr, S.C., Ruppensburg, C.C., Francis, J.W., and Katz, D.J. (2014). SPR-5 and MET-2 function cooperatively to reestablish an epigenetic ground state during passage through the germ line. *Proc. Natl. Acad. Sci. USA* 111, 9509–9514.
- Klass, M., Nguyen, P.N., and Dechavigny, A. (1983). Age-correlated changes in the DNA template in the nematode *Caenorhabditis elegans*. *Mech. Ageing Dev.* 22, 253–263.
- Koh, K.P., and Rao, A. (2013). DNA methylation and methylcytosine oxidation in cell fate decisions. *Curr. Opin. Cell Biol.* 25, 152–161.
- Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* 11, 204–220.
- Liu, N., Dai, Q., Zheng, G., He, C., Parisien, M., and Pan, T. (2015). N(6)-methyladenosine-dependent RNA structural switches regulate RNA-protein interactions. *Nature* 518, 560–564.
- Machanick, P., and Bailey, T.L. (2011). MEME-ChIP: motif analysis of large DNA datasets. *Bioinformatics* 27, 1696–1697.
- Martin, C., and Zhang, Y. (2007). Mechanisms of epigenetic inheritance. *Curr. Opin. Cell Biol.* 19, 266–272.
- Meselson, M., and Yuan, R. (1968). DNA restriction enzyme from *E. coli*. *Nature* 217, 1110–1114.
- Moazed, D. (2011). Mechanisms for the inheritance of chromatin states. *Cell* 146, 510–518.
- Morgan, H.D., Sutherland, H.G., Martin, D.I., and Whitelaw, E. (1999). Epigenetic inheritance at the agouti locus in the mouse. *Nat. Genet.* 23, 314–318.
- Nottke, A.C., Beese-Sims, S.E., Pantalena, L.F., Reinke, V., Shi, Y., and Colaiacovo, M.P. (2011). SPR-5 is a histone H3K4 demethylase with a role in meiotic double-strand break repair. *Proc. Natl. Acad. Sci. USA* 108, 12805–12810.
- Rechavi, O., Hourri-Ze'evi, L., Anava, S., Goh, W.S., Kerk, S.Y., Hannon, G.J., and Hobert, O. (2014). Starvation-induced transgenerational inheritance of small RNAs in *C. elegans*. *Cell* 158, 277–287.
- Sado, T., Fenner, M.H., Tan, S.S., Tam, P., Shioda, T., and Li, E. (2000). X inactivation in the mouse embryo deficient for Dnmt1: distinct effect of hypomethylation on imprinted and random X inactivation. *Dev. Biol.* 225, 294–303.
- Saksouk, N., Barth, T.K., Ziegler-Birling, C., Olova, N., Nowak, A., Rey, E., Mateos-Langerak, J., Urbach, S., Reik, W., Torres-Padilla, M.E., et al. (2014). Redundant mechanisms to form silent chromatin at pericentromeric regions rely on BEND3 and DNA methylation. *Mol. Cell* 56, 580–594.
- Shi, Y., Lan, F., Matson, C., Mulligan, P., Whetstone, J.R., Cole, P.A., Casero, R.A., and Shi, Y. (2004). Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* 119, 941–953.
- Simpson, V.J., Johnson, T.E., and Hammen, R.F. (1986). *Caenorhabditis elegans* DNA does not contain 5-methylcytosine at any time during development or aging. *Nucleic Acids Res.* 14, 6711–6719.
- Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* 14, 204–220.
- Stadler, M.B., Murr, R., Burger, L., Ivanek, R., Lienert, F., Schöler, A., van Nimwegen, E., Wirbelauer, C., Oakeley, E.J., Gaidatzis, D., et al. (2011). DNA-binding factors shape the mouse methylome at distal regulatory regions. *Nature* 480, 490–495.
- Stein, R., Razin, A., and Cedar, H. (1982). In vitro methylation of the hamster adenine phosphoribosyltransferase gene inhibits its expression in mouse L cells. *Proc. Natl. Acad. Sci. USA* 79, 3418–3422.
- Teixeira, F.K., Heredia, F., Sarazin, A., Roudier, F., Boccara, M., Ciaudo, C., Cruaud, C., Poulain, J., Berdasco, M., Fraga, M.F., et al. (2009). A role for RNAi in the selective correction of DNA methylation defects. *Science* 323, 1600–1604.
- Wassenegger, M., Heimes, S., Riedel, L., and Sänger, H.L. (1994). RNA-directed de novo methylation of genomic sequences in plants. *Cell* 76, 567–576.
- Wenzel, D., Palladino, F., and Jedrusik-Bode, M. (2011). Epigenetics in *C. elegans*: facts and challenges. *Genesis* 49, 647–661.
- Wion, D., and Casadesús, J. (2006). N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. *Nat. Rev. Microbiol.* 4, 183–192.
- Yi, C., and He, C. (2013). DNA repair by reversal of DNA damage. *Cold Spring Harb. Perspect. Biol.* 5, a012575.
- Zemach, A., and Zilberman, D. (2010). Evolution of eukaryotic DNA methylation and the pursuit of safer sex. *Curr. Bio.* 20, R780–R785.
- Zemach, A., McDaniel, I.E., Silva, P., and Zilberman, D. (2010). Genome-wide evolutionary analysis of eukaryotic DNA methylation. *Science* 328, 916–919.

N^6 -Methyldeoxyadenosine Marks Active Transcription Start Sites in *Chlamydomonas*

Graphical Abstract



Authors

Ye Fu, Guan-Zheng Luo, ..., Laurens Mets, Chuan He

Correspondence

chuanhe@uchicago.edu

In Brief

DNA methylation on N^6 -adenine is distributed within the *Chlamydomonas* genome in a manner distinct from the more-studied cytosine methyl marks and is associated with the transcriptional start sites of active genes.

Highlights

- Genome-wide profiling reveals a bimodal distribution of 6mA enriched around TSS
- 6mA marks active genes in *Chlamydomonas*
- A periodic pattern of 6mA at base resolution correlates with nucleosome positioning
- 6mA exclusively marks DNA linkers between adjacent nucleosomes around TSS

Accession Numbers

GSE62690



N⁶-Methyldeoxyadenosine Marks Active Transcription Start Sites in *Chlamydomonas*

Ye Fu,^{1,2,4,5} Guan-Zheng Luo,^{1,2,5} Kai Chen,^{1,2} Xin Deng,^{1,2} Miao Yu,^{1,2} Dali Han,^{1,2} Ziyang Hao,^{1,2} Jianzhao Liu,^{1,2} Xingyu Lu,^{1,2} Louis C. Doré,^{1,2} Xiaocheng Weng,^{1,2} Qianjiang Ji,^{1,2} Laurens Mets,³ and Chuan He^{1,2,*}

¹Department of Chemistry and Institute for Biophysical Dynamics, The University of Chicago, 929 East 57th Street, Chicago, IL 60637, USA

²Howard Hughes Medical Institute, The University of Chicago, 929 East 57th Street, Chicago, IL 60637, USA

³Department of Molecular Genetics and Cell Biology, The University of Chicago, 920 East 58th Street, Chicago, IL 60637, USA

⁴Present address: Department of Chemistry and Chemical Biology, Harvard University, 12 Oxford Street, Cambridge, MA 02138, USA

⁵Co-first author

*Correspondence: chuanhe@uchicago.edu

<http://dx.doi.org/10.1016/j.cell.2015.04.010>

SUMMARY

N⁶-methyldeoxyadenosine (6mA or m⁶A) is a DNA modification preserved in prokaryotes to eukaryotes. It is widespread in bacteria and functions in DNA mismatch repair, chromosome segregation, and virulence regulation. In contrast, the distribution and function of 6mA in eukaryotes have been unclear. Here, we present a comprehensive analysis of the 6mA landscape in the genome of *Chlamydomonas* using new sequencing approaches. We identified the 6mA modification in 84% of genes in *Chlamydomonas*. We found that 6mA mainly locates at ApT dinucleotides around transcription start sites (TSS) with a bimodal distribution and appears to mark active genes. A periodic pattern of 6mA deposition was also observed at base resolution, which is associated with nucleosome distribution near the TSS, suggesting a possible role in nucleosome positioning. The new genome-wide mapping of 6mA and its unique distribution in the *Chlamydomonas* genome suggest potential regulatory roles of 6mA in gene expression in eukaryotic organisms.

INTRODUCTION

Covalent modifications of individual bases in DNA can encode inheritable genetic information beyond the four canonical DNA bases (Bird, 2007). Methylations of DNA, including 5mC (Sasaki and Matsui, 2008) and 6mA (Wion and Casadesús, 2006), are the most abundant modifications in both prokaryotic and eukaryotic organisms. The well-studied 5mC modification in multicellular eukaryotes regulates diverse cellular and developmental processes (Law and Jacobsen, 2010; Smith and Meissner, 2013); however, the biological function of 6mA in eukaryotes is still unclear.

6mA is known to be present in the genomic DNA of viruses, bacteria, protists, fungi, and algae and has been detected in plant DNA and mosquito DNA (Ratel et al., 2006). In bacteria, 6mA plays crucial roles in the regulation of DNA mismatch repair

(Messer and Noyer-Weidner, 1988), chromosome replication (Lu et al., 1994), cell defense, cell-cycle regulation (Collier et al., 2007), transcription, and virulence (Low et al., 2001). The maps of 6mA in several bacteria strains have been obtained by using single-molecule real-time (SMRT) sequencing (Fang et al., 2012; Murray et al., 2012).

Besides bacteria, certain unicellular eukaryotes also contain 6mA in their genomes. For instance, the protozoan *Tetrahymena* (Hattman et al., 1978), *Oxytricha fallox* (Rae and Spear, 1978), and *Paramecium aurelia* (Cummings et al., 1974) have relatively abundant 6mA but little 5mC. On the other hand, green algae *Chlamydomonas reinhardtii* (Hattman et al., 1978) and *Volvox carteri* (Babinger et al., 2001) possess both 6mA and 5mC. Although common in bacteria, no corresponding restriction endonucleases have been reported in these species. Therefore, 6mA in these unicellular eukaryotic genomes has long been suspected of possessing functions other than exclusion of foreign DNA or viruses (Ehrlich and Zhang, 1990). Additionally, evidences for the existence of 6mA in plants, insects, and mammals have also been reported.

Chlamydomonas reinhardtii (referred to hereafter as *Chlamydomonas*) is a unicellular green alga that has been widely used as a model organism to study photosynthesis, eukaryotic flagella, and biomass production (Merchant et al., 2007). The high level (~0.3–0.5 mol%) of 6mA in the nuclear DNA of *Chlamydomonas* (Hattman et al., 1978) prompted us to study its distribution and function, which could help to decipher the long mystery of 6mA in eukaryotes and to develop bioengineering tools that may facilitate biomass and biofuel production (Radakovits et al., 2010).

In this study, we employed/developed several methods for mapping 6mA sites in genomic DNA. We first applied 6mA immunoprecipitation sequencing, or 6mA-IP-seq, which is an antibody-based profiling method to obtain the genome-wide distribution of 6mA. We then developed a 6mA-CLIP-exo strategy of employing photo-crosslinking followed by exonuclease digestion to achieve a much higher resolution. Lastly, we developed a restriction enzyme-based 6mA sequencing, or 6mA-RE-seq, to detect 6mA sites at single-nucleotide resolution in genome wide. Application of these three approaches to the *Chlamydomonas* genome revealed that 6mA marks more than 14,000 genes, accounting for 84% of all *Chlamydomonas* genes. This

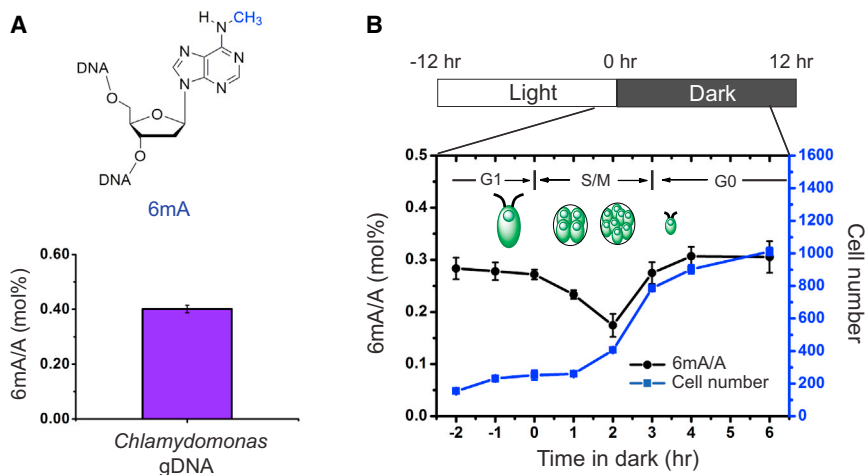


Figure 1. The Presence and Conservation of 6mA in *Chlamydomonas* Genomic DNA

(A) The presence of 6mA in isolated *Chlamydomonas* genomic DNA is determined by UHPLC-QQQ-MS/MS. Ratios of 6mA/A are shown ($n = 6$, mean \pm SEM).

(B) The level of 6mA decreases at the beginning of S/M phase during DNA replication and increases back to the original level at the late stage of S/M phase. During a multiple fission cell cycle in naturally synchronized cells, cells begin to replicate the genomic DNA 1 hr after dark and divide two times, leading to ~ 4 -fold increase of the total cell number. Ratios of 6mA/A are shown on the left axis ($n = 4$, mean \pm SEM) with calibrated cell concentration (cells/ μ l) shown on the right axis ($n = 2$, mean \pm SEM).

See also Figure S1.

methylation is highly enriched around transcription start sites (TSS) with a bimodal distribution and significant local depletion at TSS. We used RNA-seq to quantify gene expression and found that the presence of 6mA is correlated with actively expressed genes. This pattern is distinct from that of 5mC, which accumulates mostly in gene bodies in *Chlamydomonas*. At single-nucleotide resolution, we also discovered that 6mA is enriched around TSS but exhibits an unexpected, strongly periodic pattern, suggesting controlled deposition of 6mA in association with nucleosome spacing. Nucleosome profiling revealed that 6mA around TSS occurs primarily within the linker DNA between nucleosomes. Our data show that 6mA is an abundant DNA mark associated with actively expressed genes in *Chlamydomonas*. These methods and results should stimulate future functional investigations of 6mA in *Chlamydomonas* and other eukaryotic organisms.

RESULTS

6mA Is a Stable Modification in *Chlamydomonas* Genomic DNA

To accurately quantify the level of 6mA in genomic DNA, we applied an LC-MS/MS assay using pure 6mA nucleoside as an external standard (Figures S1A and S1B) (Jia et al., 2011). In agreement with the previous data (Hattman et al., 1978), we detected ~ 0.4 mol% of 6mA (6mA/A) in the genomic DNA isolated from *Chlamydomonas* cultured in mixotrophic conditions, i.e., Tris-Acetate-Phosphate (TAP) medium under constant light (Figure 1A).

To determine whether 6mA is stable during cell growth, we monitored the 6mA level during a multiple fission cell cycle in naturally synchronized cells induced by a 12 hr/12 hr light/dark cycle in minimal media cultures (Bisova et al., 2005). Under such growth conditions, cells grow in size during the light phase (G1 phase) and then undergo two to three rapid rounds of alternating DNA replications and cellular divisions (S/M phase) from 1 hr to 5 hr after entering the dark phase. Cells were mostly synchronized and rapidly divided under this light-dark phase transition according to cell counting measured by flow cytometry (Figure 1B). The proportion of 6mA in genomic DNA was measured

before and after the switch from light to dark. Our results showed that the overall 6mA level in genomic DNA decreased by $\sim 40\%$ in 2 hr after dark, corresponding to the time period when DNA was replicated. This level then rose quickly back to the original level within 2 hr. This result indicated that 6mA is installed on the newly synthesized DNA within a short time period after DNA replication and is stably maintained during cell proliferation (Figure 1B).

Genome-wide Mapping of 6mA with 6mA-IP-Seq

Although the existence of 6mA in *Chlamydomonas* has been known, its distribution/localizations are unclear. To generate a de novo map of the genome-wide distribution of 6mA, we applied 6mA-IP-seq. Similar to the methylated DNA immunoprecipitation (MeDIP) (Weber et al., 2005) that has been widely applied to enrich 5mC-containing DNA fragments, we sought to use a 6mA-specific antibody to enrich the 6mA-containing DNA fragments. An antibody that recognizes the N⁶-methyladenine base has recently been applied to genome-wide profiling of 6mA sites in RNA (Dominissini et al., 2012; Meyer et al., 2012). By performing dot-blot assay on synthesized 6mA-containing DNA oligonucleotide, we confirmed that this anti-6mA antibody can also specifically recognize 6mA in both single-stranded and double-stranded DNA (Figure S2).

We then isolated genomic DNA from *Chlamydomonas* and fragmented it into 200–400 base pairs by sonication. The fragmented DNA was ligated to an adaptor with specific index sequence (Figure 2), which was then denatured to single-stranded DNA, and immunoprecipitated using the anti-6mA antibody. The captured DNA was eluted through the competition with 6mA single nucleotide and PCR amplified to construct the DNA library (Figure 2). Simultaneously, an input library was obtained by PCR amplification of the ligated DNA before immunoprecipitation. Both libraries were subjected to high-throughput sequencing. The obtained sequencing reads were mapped to a reference genome of *Chlamydomonas* (JGI version 9.1), and 6mA sites were identified using a peak-detection algorithm (Zhang et al., 2008). The false detection rate (FDR) was estimated to be below 0.01.

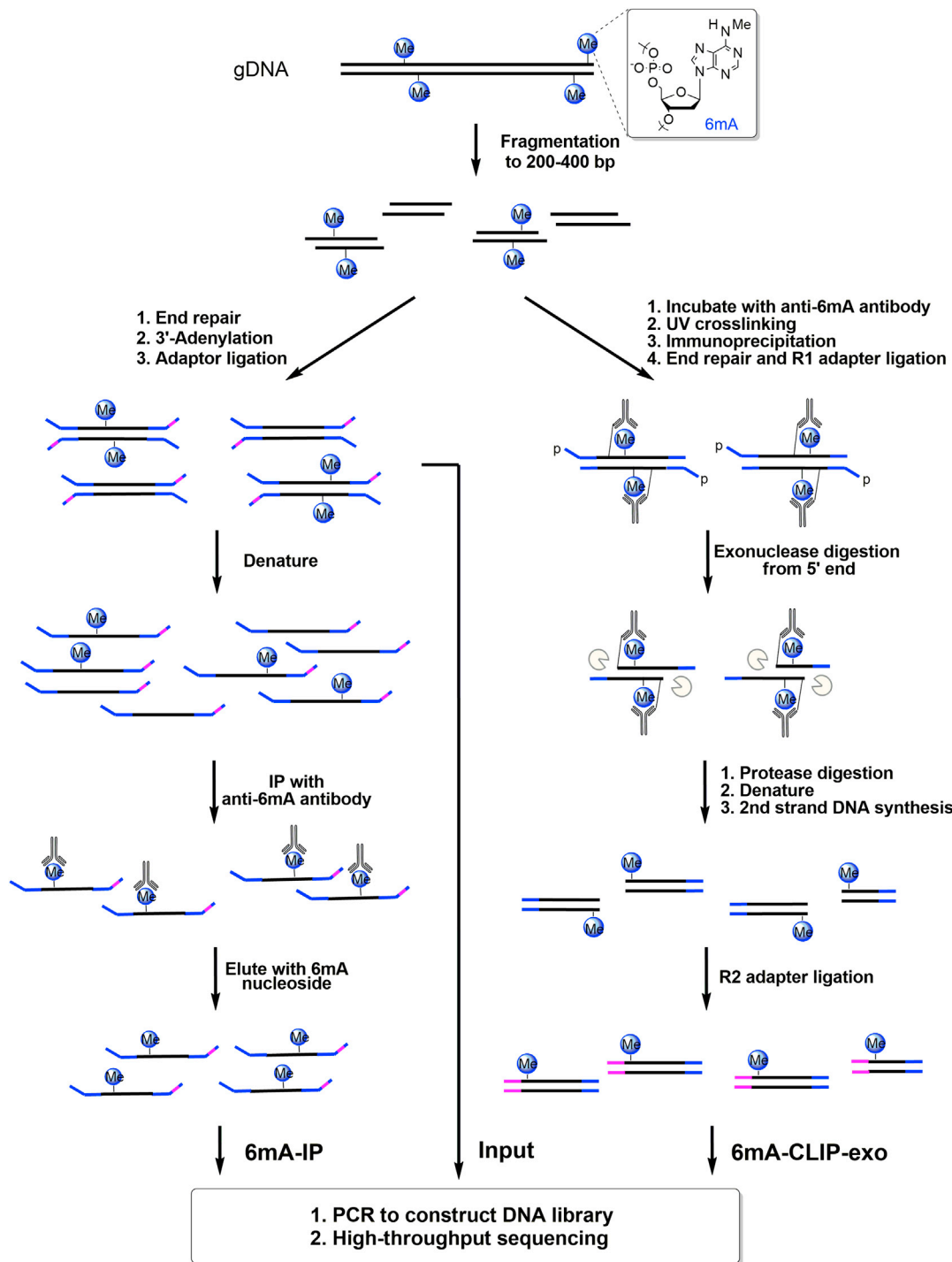


Figure 2. Schematic Diagram of 6mA-IP-Seq and 6mA-CLIP-Exo

For 6mA-IP-seq (left), fragmented genomic DNA (gDNA) is ligated to a Y-shaped adaptor with specific index sequence, denatured, and immunoprecipitated using anti-6mA antibody. The captured DNA is eluted with 6mA single nucleotide and PCR amplified to construct the DNA library. Simultaneously, the input library was obtained from the ligated DNA before immunoprecipitation. For 6mA-CLIP-exo (right), fragmented gDNA is incubated with anti-6mA antibody, crosslinked by 254 nm UV irradiation, and immunoprecipitated. The crosslinked DNA is ligated to adaptor R1 on beads, followed by 5' to 3' exonuclease digestion. Antibody-protected DNA is preserved, and a 2nd-strand DNA synthesis is performed after protease digestion of the antibody. A second ligation to adaptor R2 provides the template for PCR amplification to construct the library for high-throughput sequencing. Boundaries were determined by the sequencing ends of the 6mA-CLIP-exo-seq to provide a high-resolution localization of 6mA.

See also [Figure S2](#).

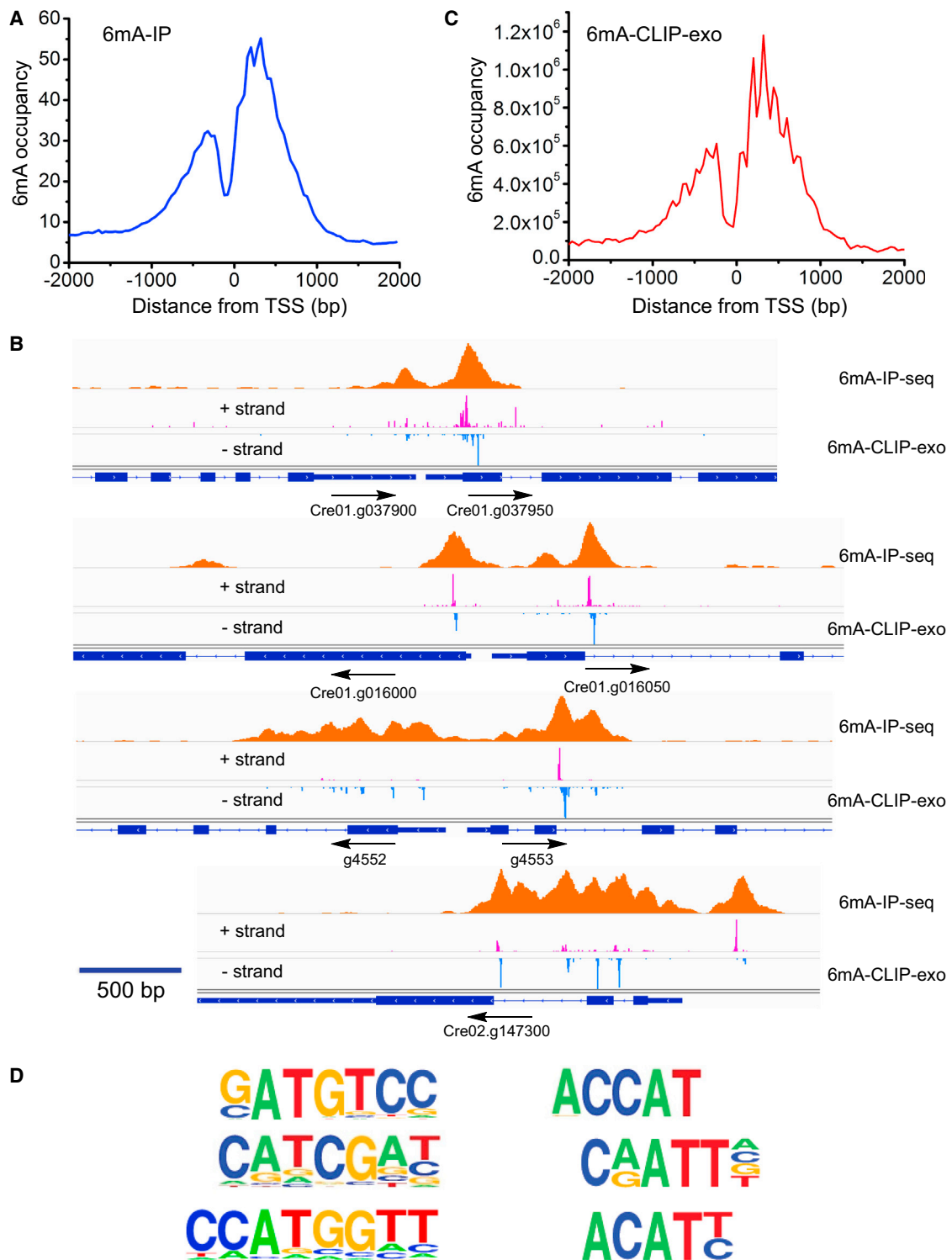


Figure 3. A Bimodal Distribution of 6mA around Transcription Start Sites

(A) Distribution of 6mA peaks around TSS measured by 6mA-IP-seq. 6mA is enriched around TSS with a bimodal distribution and a local depletion at TSS. 6mA occupancy represents the reads coverage averaged by gene number in 6mA-IP-seq.

(B) Snapshot of 6mA peak determined by both 6mA-IP-seq and 6mA-CLIP-exo in specific gene loci. 6mA peaks can be detected both upstream and downstream of TSS in single direction promoter region and bidirectional promoter region. Some enrichment peaks are located in the first and second introns. Boundaries of

(legend continued on next page)

6mA Bases Are Highly Enriched around TSS with a Bimodal Distribution

We performed 6mA-IP-seq on *Chlamydomonas* cultured under mixotrophic (constant light) or heterotrophic (constant dark) conditions in TAP medium during the pre-stationary phase. For each condition, we performed two biological replicates. After peak calling, we identified 25,803 and 28,982 high-confidence 6mA peaks in light samples and 22,005 and 21,016 peaks in dark samples (FDR < 0.01), respectively. Among them, more than 95% of the peaks mutually occur in both replicate samples, indicating the high reproducibility of our approach (Figure S3A). About 88% of the peaks are common under both light and dark conditions, suggesting a faithful installation/maintenance mechanism of 6mA at specific genomic regions. Consistent with the previous measurements that 6mA was only detected in *Chlamydomonas* nuclear DNA but not chloroplast DNA, all the 6mA peaks were mapped to the nuclear genome but not the chloroplast genome. To our surprise, we observed that 6mA is highly enriched around the TSS of 14,868 genes, constituting 84% of all the genes in the *Chlamydomonas* genome (Figure 3A). A closer examination of the distribution revealed that the 6mA sites enriched around TSS (−500 to +800 bp, ~91% of all 6mA peaks) exhibit a bimodal distribution with a significant local depletion at TSS. The summit of the peak tends to locate within 500 bp downstream of TSS (Figure 3A). The rest of the 6mA peaks (~9%) not associated with TSS do not show specific patterns and reside in both gene bodies and intergenic regions. The average peak width of the identified peaks is around 320 bp, which is consistent with the fragmentation size of our sequenced DNA (200–400 bp). We cannot quantify the number of methylation sites under each 6mA peak; however, some peaks are noticeably broader, with certain peaks containing multiple sub-peak summits, suggesting the presence of multiple methylation sites in these regions (Figure 3B). Thus, our observation revealed a region-specific bimodal methylation pattern of 6mA highly enriched around TSS.

6mA-CLIP-Exo with Immunoprecipitation, Photo-Crosslinking, and Exonuclease Digestion

Inspired by chromatin immunoprecipitation followed by exonuclease digestion (ChIP-exo), a method to map the locations at which a protein binds to the genome (Rhee and Pugh, 2012), we introduced photo-crosslinking after the antibody-based 6mA enrichment (Chen et al., 2015) followed by exonuclease digestion in an attempt to identify 6mA peaks with higher resolution. DNA/antibody complexes were covalently crosslinked with UV irradiation before being captured by magnetic Protein A beads. The crosslinked DNA was ligated to adaptor R1 before being treated with two 5′-3′ exonucleases, Lambda exonuclease and RecJ_f exonuclease, to digest the DNA from the 5′ end. The presence of crosslinked antibody stopped

the exonuclease digestion before the crosslinking site. Antibody was then removed by proteinase K digestion, and DNA fragments were recovered for primer extension. The double-stranded DNA (dsDNA) product was ligated to adaptor R2 and sequenced (Figure 2). By mapping the read ends to the *Chlamydomonas* genome, we determined the boundary sites of antibody-protected regions, which contain one or more 6mA sites. As expected, we successfully improved the resolution to ~33 bp (Figure 3B) and identified 30,899 6mA-containing sequences with 67% overlapping with 6mA peaks identified from 6mA-IP-seq. Meanwhile, 73% of 6mA peaks from 6mA-IP-seq contain at least one 6mA-containing sequence identified from 6mA-CLIP-exo (Figure S3B). These higher-resolution 6mA peaks showed the same enrichment around TSS with a bimodal distribution, a local depletion at TSS, and a potential periodic pattern (Figures 3C and S3C). A motif search revealed multiple high-frequency sequences (Figure 3D), most of which contain an ApT dinucleotide motif (Figure 3D), reminiscent of the CpG methylation in most eukaryotic organisms and suggesting ApT as the general consensus sequence.

Validation of Individual Methylation Sites

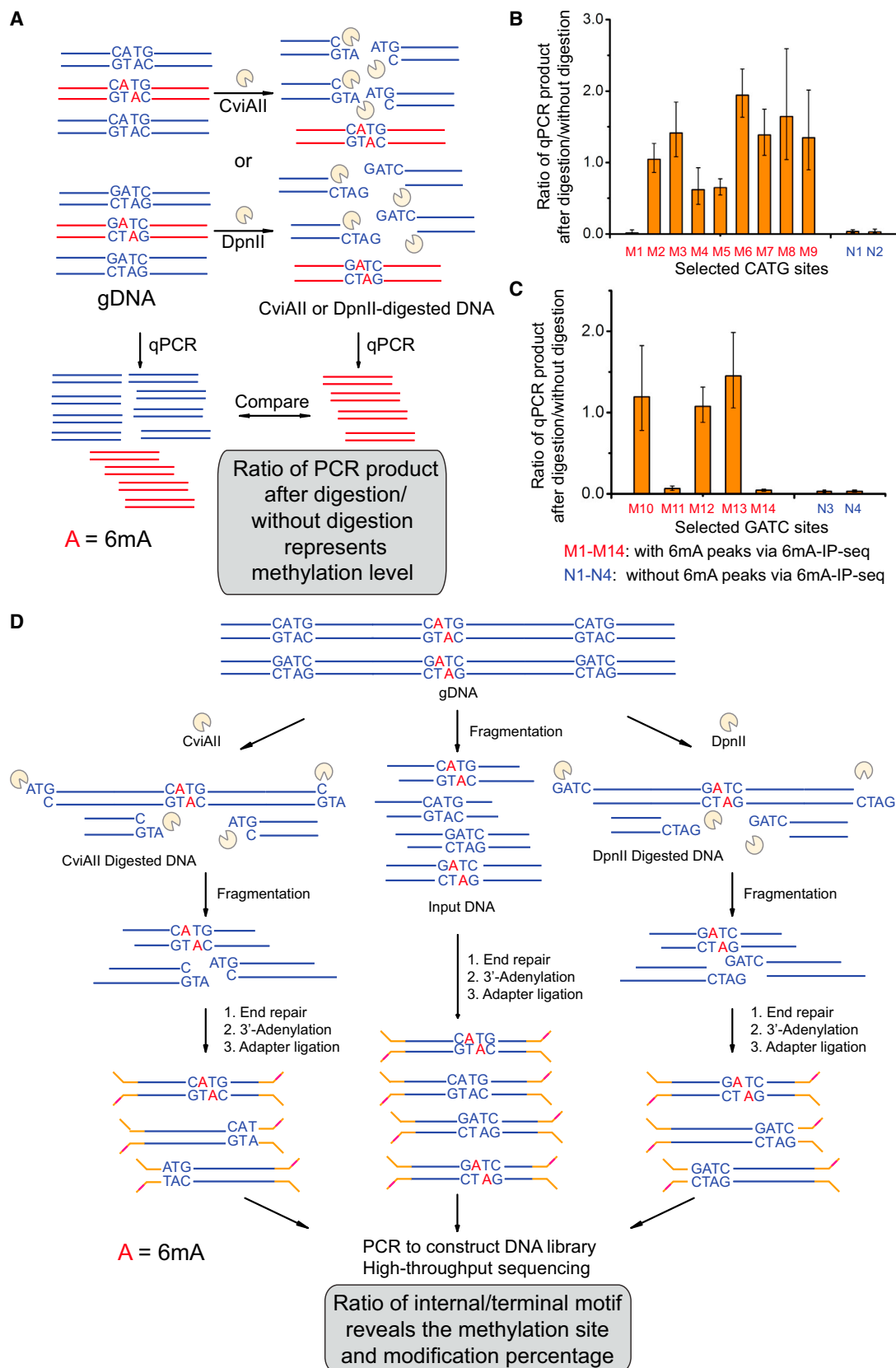
The methylation status of 6mA in specific motif sites can be validated by digestion with restriction enzymes originating from bacteria and viruses that are sensitive to 6mA methylation. For instance, CviAII is sensitive to 6mA and only digests the unmethylated CATG sequence (Zhang et al., 1992), whereas DpnII cuts only the unmethylated GATC sequence (Vovis and Lacks, 1977). We then applied the restriction-enzyme-digestion assay followed by quantitative PCR (6mA-RE-qPCR) to quantitatively evaluate the methylation status on specific motif sequences (Figure 4A). In this assay, we treated the isolated genomic DNA with CviAII or DpnII overnight to fully digest the unmethylated recognition motifs. We then designed PCR primers to specifically amplify the region flanked by the candidate 6mA site. In principle, the percentage of 6mA in the target 6mA site could be determined by quantitative PCR (qPCR) amplification of the restriction enzyme digested genomic DNA using undigested genomic DNA as a control, given that 6mA hinders digestion. This strategy was tested by analyzing nine specific CATG sites and five specific GATC sites within identified 6mA peaks from 6mA-IP-seq and 6mA-CLIP-exo, along with two CATG sites and two GATC sites in regions that are not methylated based on 6mA-IP-seq results. The 6mA-RE-qPCR assay identified 8/9 of these CATG sites and 3/5 GATC sites within 6mA peaks to be completely or partially (a lower methylation frequency in a population of DNA molecules) methylated. The control sites not identified by 6mA-IP-seq were not methylated by this assay (Figures 4B and 4C). Therefore, this assay provides locus-specific validation of the 6mA-IP-seq results.

6mA-CLIP-exo-seq on both DNA strands were marked by magenta and blue color. Regions between the two nearest boundaries were determined as a 6mA-containing sequence. Black arrows indicate the transcription direction.

(C) Distribution of 6mA peaks around TSS measured by 6mA-CLIP-exo. The enrichment of 6mA near TSS shows a similar pattern as that obtained using 6mA-IP-seq. In addition, several spikes could be observed from the large peak.

(D) The dinucleotide sequence ApT is enriched in 6mA-CLIP-exo peaks, including CATG.

See also Figure S3.



(legend on next page)

Genome-wide Identification of Single 6mA Sites Using 6mA-RE-Seq

The *Chlamydomonas* genome is GC rich (G+C content 64%) and ~120 million base pairs (Merchant et al., 2007). The ~0.4 mol% 6mA/A ratio corresponds to ~85,000 fully methylated 6mA sites. Our 6mA-IP-seq identified roughly 25,000 peaks, with each peak potentially covering multiple 6mA sites (most of them are fully methylated at ~100%, see below), consistent with 6mA-IP-seq results showing that most 6mA peaks in the *Chlamydomonas* genome cluster around TSS sites. The 6mA-CLIP-exo results revealed several high-frequency sequences that include CATG and GATC. After we validated these two sequences as genuine 6mA methylation sites that mark TSS regions in *Chlamydomonas*, we sought to develop a high-throughput assay to map 6mA methylation in these selected sequences in genome wide at single-base resolution and to quantitatively determine the modification percentage at each site.

Genomic DNA was isolated and treated with CviAII or DpnII and then sonicated to ~300 base pair fragments, end-repaired by T4 DNA polymerase, 3'-adenylated, and ligated to DNA adapters. The unmethylated CATG or GATC motifs would be digested and should be enriched at the end of the DNA fragments. The methylated motifs should resist restriction enzyme-mediated digestion and be present in the internal locations of DNA fragments. After PCR amplification of the fragments, a DNA library can be prepared for high-throughput sequencing. The ratio of a specific CATG or GATC sites with sequence reads internal versus at the end represents the relative methylation to unmethylation ratio. An input sample from genomic DNA without enzyme digestion serves as a control. Through mapping sequencing reads to the reference genome, we can identify the methylation status for every CATG or GATC motif in genome wide. We named this approach—as diagrammed in Figure 4D—6mA-RE-seq and applied it to *Chlamydomonas* genomic DNA. While the specificity of DpnII to non-methylated DNA has been well characterized (Vovis and Lacks, 1977), the specificity of CviAII in cutting only non-methylated but not hemi- or fully methylated sequences was further confirmed using synthetic DNA probes (Figure S4A).

By applying 6mA-RE-seq to two biologically independent samples of *Chlamydomonas* grown under constant light or dark conditions, we obtained a high-resolution 6mA map of all CATG and GATC motifs in the *Chlamydomonas* genome. As expected, most of the sequencing reads were initiated with ATG or GATC for samples digested by CviAII or DpnII, which resulted from the digestion of unmethylated CATG or GATC sites,

respectively (Figures S4B and S4C). Meanwhile, the intact CATG or GATC motifs that appear internal to the sequencing reads were counted as specific 6mA sites. We developed a bioinformatics algorithm with which to calculate the methylation level of individual 6mA sites within corresponding genomic sequences by calculating the ratio of reads obtained from fragment terminals to total reads of each site. We successfully identified 24,970 and 19,778 C6mATG sites with high confidence (FDR < 0.01) in light and dark samples, respectively. 4,967 and 4,174 high-confidence G6mATC sites were found in the same samples. Among the methylated sites discovered, 15,883 C6mATG sites and 3,337 G6mATC sites were identified from both light and dark samples, showing consistency of the method and reinforcing 6mA as a persistent DNA modification in *Chlamydomonas* (Figure 5A). These single 6mA sites include methylation sites that we have also validated using 6mA-RE-qPCR (Figures 4B and 4C and Table S1). The sites without methylation based on 6mA-IP-seq and 6mA-RE-qPCR results were determined to be unmethylated by 6mA-RE-seq as well (Figures 4B and 4C and Table S1). Approximately 78% (13,076/15,883 for C6mATG and 2,069/3,337 for G6mATC) of the total detected sites overlap with 6mA peaks identified by 6mA-IP-seq (Figure 5B). We plotted base-resolution 6mA sites that overlap with corresponding 6mA peaks as identified from 6mA-IP-seq. The 6mA peaks are highly enriched around the identified single 6mA sites, with peak summits right on top of the single 6mA sites (Figure 5C). In addition, most of these methylation sites are close to 100% methylated, as indicated by the ratio of internal versus terminal sequencing reads (Figures S5A and S5B).

We performed an extended motif search based on the newly identified sites to examine whether there is any additional preference of nucleotides flanking the CATG or GATC sequence; however, no additional consensus nucleotides were observed (Figure S5C). Considering the high frequency of CATG and GATC all over the genome (588,209 CATGs and 144,087 GATCs), the methylated sites occupy only 3%–4% of all available motifs. However, the identified CATG and GATC methylations represent ~30% (24,970/85,000) and ~6% (4,967/85,000) of all genomic 6mA sites, respectively. On the other hand, there are ~28% of the 6mA-IP-seq peaks that do not contain any CATG or GATC sequences along the entire genomic regions, indicating the presence of other 6mA sites in distinct sequence contexts besides these two motifs. Interestingly, individual 6mA sites located at these two different sequence contexts tend to cluster in short regions (Figure S5D). We also

Figure 4. Single-Site Detection of 6mA Using Methylation-Sensitive Restriction Enzymes

(A) Schematic diagram of 6mA-RE-qPCR for validation of specific 6mA. Restriction enzymes CviAII or DpnII that are sensitive to 6mA methylation in CATG or GATC were used to digest the unmethylated CATG or GATC sites in genomic DNA, respectively. The undigested CATG or GATC sites represent the methylated fraction and can be PCR amplified by using primers that cover these sites.

(B and C) qPCR results of 11 selected CATG sites and 7 GATC sites validated the accuracy of 6mA-IP-seq. After CviAII- or DpnII-mediated digestion, qPCR was performed using specific primers covering these sites. Relative abundances of undigested CATG or GATC sites were calculated from the Δ Ct value between digested and undigested DNA samples ($n = 3$, mean \pm SEM).

(D) Schematic diagram of 6mA-RE-seq. gDNA is digested with CviAII or DpnII, sonicated to small fragments around 100 base pair, and constructed into sequencing libraries. The ratio for CATG or GATC internal of sequence reads versus at the end of sequence reads of a specific genomic site represents the relative methylation to unmethylation ratio. An input sample from gDNA without CviAII- or DpnII-based digestion serves as a control.

See also Figure S4 and Table S1.

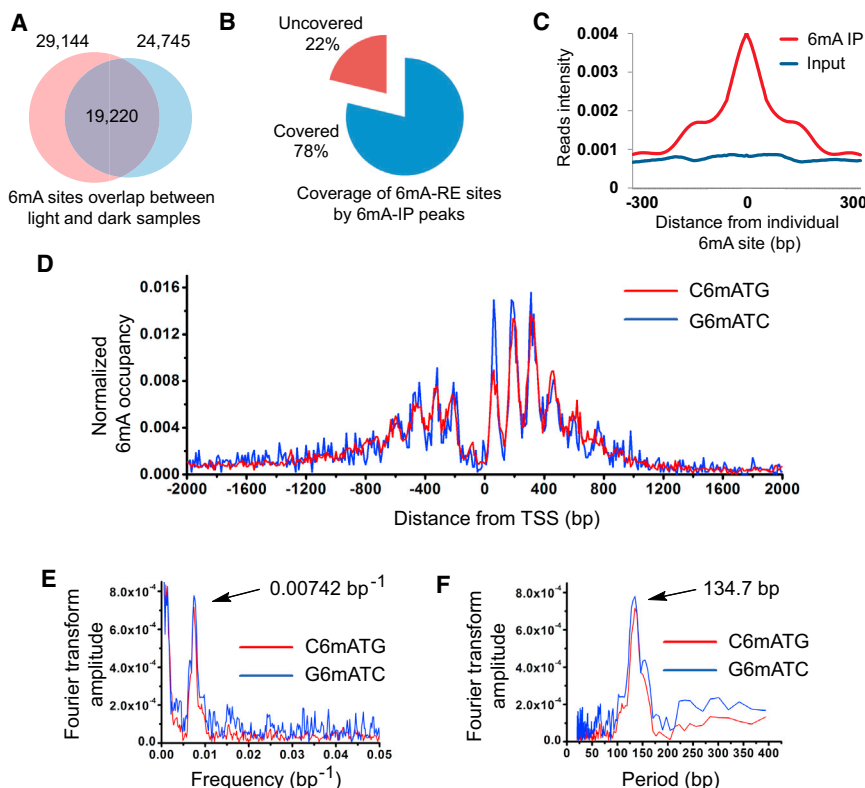


Figure 5. Single-Nucleotide-Resolution Map of 6mA

(A) Overlap of two 6mA-RE-seq samples under light and dark growth conditions. The majority of methylation sites were detected in both samples, indicating the consistency of this method. (B) A majority of the detected single 6mA sites by 6mA-RE-seq are covered by 6mA peaks identified by 6mA-IP-seq. (C) Overlap of 6mA sites identified by 6mA-RE-seq with the 6mA peak identified by 6mA-IP-seq. (D) 6mA occupancy around TSS normalized to the CATG and GATC distribution. A periodic pattern of 6mA around TSS could be observed for both C6mATG and G6mATC motifs. (E) Fourier transformation of 6mA distribution peaks. (F) Periods of the corresponding frequency in Fourier transformation. The dominant period length is 134.7 bp. See also [Figure S5](#).

observed multiple CATG and GATC motifs in a single peak identified from 6mA-IP-seq, and the peak length linearly correlates with the number of CATG or GATC motifs present in the region ([Figure S5E](#)). Taken together, these results indicate that 6mA methylation occurs mainly to ApT in multiple sequence motifs that tend to cluster together.

Periodic Distribution of 6mA near TSS Sites

To further understand the methylation specificity, we calculated the density of individual fully methylated 6mA sites around TSS (over 90% 6mA sites are close to fully methylated). Strikingly, we observed an apparent periodic pattern of 6mA distribution near the TSS region ([Figure 5D](#)). To rule out the possibility that a biased distribution of the CATG or GATC sequences caused the periodic distribution pattern, we normalized the 6mA site frequency according to motif occurrence within each region ([Figure S5F](#)). Of particular note is an obvious discontinuity between peaks upstream and downstream of TSS, which corresponds to a local depletion at TSS ([Figure 5D](#)). Fourier analysis of the periodic profile showed that the frequency is one per 130–140 bp for both downstream and upstream 6mA peaks ([Figures 5E and 5F](#)). The observed periodic pattern is similar to the one observed in the 6mA-CLIP-exo result, which is independent of sequence bias ([Figure S3C](#)). The pattern is also conserved in both biologically independent samples and is independent of culture conditions. Both motif sequences show exactly the same pattern ([Figure S5G](#)). For comparison with the fully methylated sites, we also analyzed the distribution of partially methylated sites (< 60% methylated

measured by 6mA-RE-seq, corresponding to less than 10% of all 6mA sites). These partially methylated sites are evenly distributed without any obvious pattern or periodicity ([Figure S5H](#)). It is possible that the occurrence of these sites is governed by different mechanisms than those associated with the periodic, peri-TSS sites in *Chlamydomonas*.

6mA Preferentially Locates at Linker DNA between Two Adjacent Nucleosomes

The periodic distribution pattern of 6mA around TSS prompted us to study its correlation with nucleosome positioning. We performed nucleosome footprinting, followed by high-throughput sequencing ([Chodavarapu et al., 2010](#)), to reveal the exact position of each nucleosome in the *Chlamydomonas* genome. Briefly, micrococcal nuclease (MNase) was used to digest unprotected DNA between nucleosomes while leaving the nucleosome-occupied DNA intact; the intact DNA was then subjected to library preparation and high-throughput sequencing. After MNase digestion, the purified DNA showed a clear band with ~150 bp length; the DNA is composed of the nucleosome-protected segments ([Figure S6A](#)). These DNA segments were sequenced by paired-end sequencing. The length distribution is enriched around 147 bp ([Figure S6B](#)), which perfectly matches the reported value for *Chlamydomonas* ([Lodha and Schroda, 2005](#)). When we mapped the nucleosomes and 6mA locations to the *Chlamydomonas* genome, we found that most of the 6mA sites locate between two adjacent nucleosomes ([Figure 6A](#)). We then analyzed the statistical distribution of nucleosomes relative to individual 6mA sites, which revealed that the peaks of the closest nucleosomes are enriched ~75 bp upstream and ~78 bp downstream of the 6mA sites ([Figure 6B](#)). This pattern further supports that 6mA is mostly present in regions corresponding to the linker DNA between two adjacent

nucleosomes (Figure 6C). The analysis of nucleosome-6mA correlation also showed that the downstream nucleosomes possess a progression with a steady phase of 170–180 bp periodicity (Figure S6C), whereas the upstream nucleosomes are relatively loosely phased, and this tight periodicity disappears around 2 to 3 nucleosomes away from the 6mA site.

6mA May Contribute to the Positioning of Nucleosomes in *Chlamydomonas*

To further understand the relationship between nucleosome distribution and 6mA, we plotted their density around TSS. We found that the periodic pattern of average nucleosome occupancy around TSS in *Chlamydomonas* has distinct features compared to other species (Figure 6D): first, the density of nucleosomes around TSS is much lower than that in gene body regions and upstream promoter regions; second, the periodicity between two nucleosomes is centered at 183 bp upstream of TSS but has multiple period values downstream of TSS, including 171, 151, and 128 bp (Figure S6D). Previous studies of nucleosome distribution in *Chlamydomonas* and other organisms revealed that nucleosome-depleted regions (NDRs) are, on average, ~155–160 bp around TSS, and nucleosomes downstream of the NDRs are strictly phased in a 165–185 bp period, depending on the length of linker region between two adjacent nucleosomes (Huff and Zilberman, 2014; Lodha and Schroda, 2005). The multiple periodic values we observed could be a result of convolution between the regular nucleosome periodicity of ~170 bp and the 6mA-influenced periodicity of 130–140 bp downstream of TSS on DNA. Nonetheless, when we compared the nucleosome distribution with the 6mA distribution around TSS, we found that they correlated with each other with ~180 degree phase shift, which is consistent with our finding that 6mA preferentially locates at linker regions. To probe the relationship of 6mA distribution and nucleosome positioning in detail, we divided all the genes into two groups: with or without 6mA around TSS. Interestingly, nucleosomes phase well for genes that contain 6mA around TSS, whereas the nucleosome phase pattern was weak for genes without 6mA (Figures 6E and S6E). Taking these results together, we propose a model in which the DNA 6mA modification either restricts or marks the positions of nucleosomes near TSS in *Chlamydomonas* (Figure 6F). The 130–140 periodic pattern of 6mA leads to out-of-phase distribution and partial occupancy of nucleosomes around TSS. For example, if the distance between two adjacent 6mA sites is larger than the length of a nucleosome, such as 270 bp, one nucleosome may reside between two adjacent 6mA, in place depending on the sequence content. If the distance between two adjacent 6mA sites is shorter than 150 bp, such as 135 bp, nucleosome will be missing, leaving a nucleosome-free region between them (Figures 6A and 6F). The distribution pattern of 6mA may restrict the pattern of nucleosome positioning for each gene, such that the genome-wide pattern of nucleosome is correlated with 6mA distribution pattern.

6mA Marks the TSS Regions of Actively Transcribed Genes

The bimodal localization of 6mA around TSS prompted us to investigate its relationship with gene expression. We used

RNA-seq to analyze the expression of individual genes. We divided genes into two groups: high expression (80% of all genes) and low expression (20%) and plotted their 6mA peak abundances obtained from 6mA-CLIP-exo experiments (Figure 7A). We found a general trend that genes with lower expression tend to have low occupancies of 6mA around TSS regions. Specifically, among the 16% of genes without 6mA, ~64% are categorized as low expression or non-active genes. Correspondingly, on a genome-wide level, genes with 6mA around TSS express significantly higher than genes without 6mA (Figure S7A). The widely studied 5mC methylation typically plays repressive roles in the regulation of gene expression. However, our results reveal that 6mA marks the TSS regions of actively transcribed genes in *Chlamydomonas*. Studies have shown that 6mA can reduce the stability of the DNA duplex due to the requirement of unfavorable *trans*- configuration for base pairing. The presence of 6mA may lower the energy required for opening up the DNA duplex (Engel and von Hippel, 1978). Based on the observed periodic distribution pattern, the tightly controlled deposition of 6mA is associated with nucleosome phasing around TSS. These 6mA modifications could affect nucleosome positioning or recruit protein factors analogous to methyl-CpG-binding proteins as potential “readers” to impact transcription initiation (Sternberg, 1985). Indeed, barley nuclear extract has been shown to contain specific 6mA-binding proteins, and 6mA embedded within GATC at the promoter region can increase the transcription activity of a transfected plasmid (Rogers and Rogers, 1995).

To study potential effects of 6mA on gene regulation, we profiled the mRNA transcriptome of algae cultured under constant light and dark conditions and found 4,866 differentially expressed genes. In parallel, we used the restriction enzyme-based method to quantify the methylation level of individual 6mA site under light and dark conditions. 6mA levels in most genes were similar under both light and dark conditions (Figure S7B). These results suggest that 6mA is a general mark of TSS regions that could be actively transcribed. Transcription factors and other factors may play more direct roles in determining the exact expression levels of individual genes.

6mA and 5mC Mark Distinct Regions in the *Chlamydomonas* Genome

As 5mC is also present in high abundance in the *Chlamydomonas* genome, we wondered if any relationship exists between these two DNA base modifications. *Chlamydomonas* has an unusual pattern of 5mC methylation—overall, it has less CpG methylation compared to multicellular eukaryotes but possesses all three types of methylation of CpG, CHG, and CHH enriched in exons of genes and has only CpG methylation enriched in repeats and transposons (Feng et al., 2010). We compared bisulfite sequencing data of 5mC with the 6mA distribution that we generated. There is no specific enrichment pattern of 5mC distribution around TSS regions (Figure S7C), and 5mC generally do not co-localize with 6mA (Figure 7B). 5mC appears mostly in gene bodies with a much broader distribution and is absent near TSS regions (Figures 7C and S7C). In addition, 5mC has been proposed to be negatively

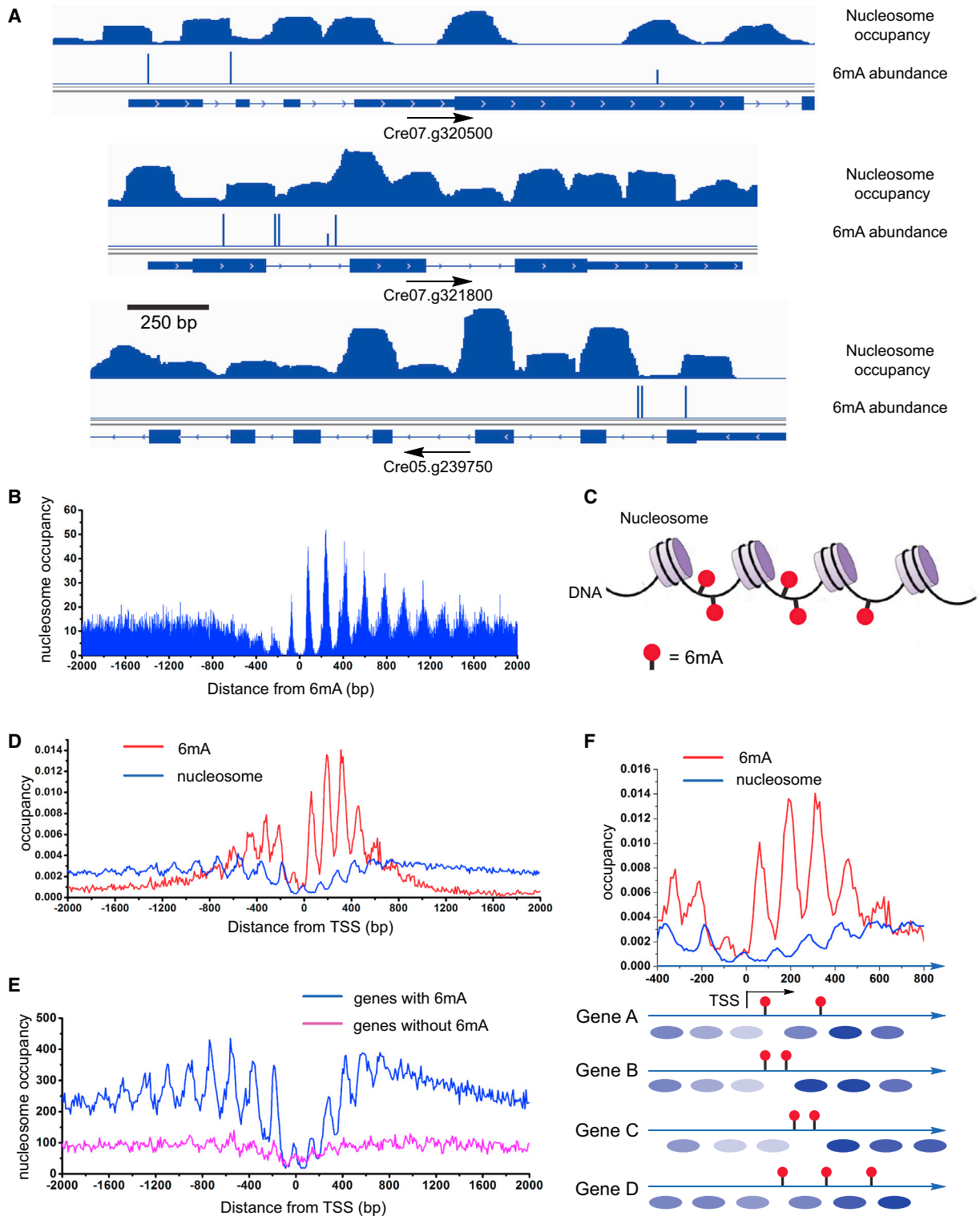


Figure 6. 6mA Resides at the DNA Linker Region between Adjacent Nucleosomes

(A) Distribution of nucleosome and 6mA in selected genes. 6mA mainly lies at the boundary region of nucleosomes. Nucleosome occupancy is shown on the first line, and 6mA sites identified from 6mA-RE-seq are shown on the second line. Genome annotations are shown on the bottom line.

(B) Nucleosome occupancy around 6mA sites. 0 defines the 6mA site, with downstream noted as positive. Nucleosomes reside adjacent to but not on the 6mA site. Nucleosomes downstream of 6mA sites show a constant period of ~170–180 bp.

(legend continued on next page)

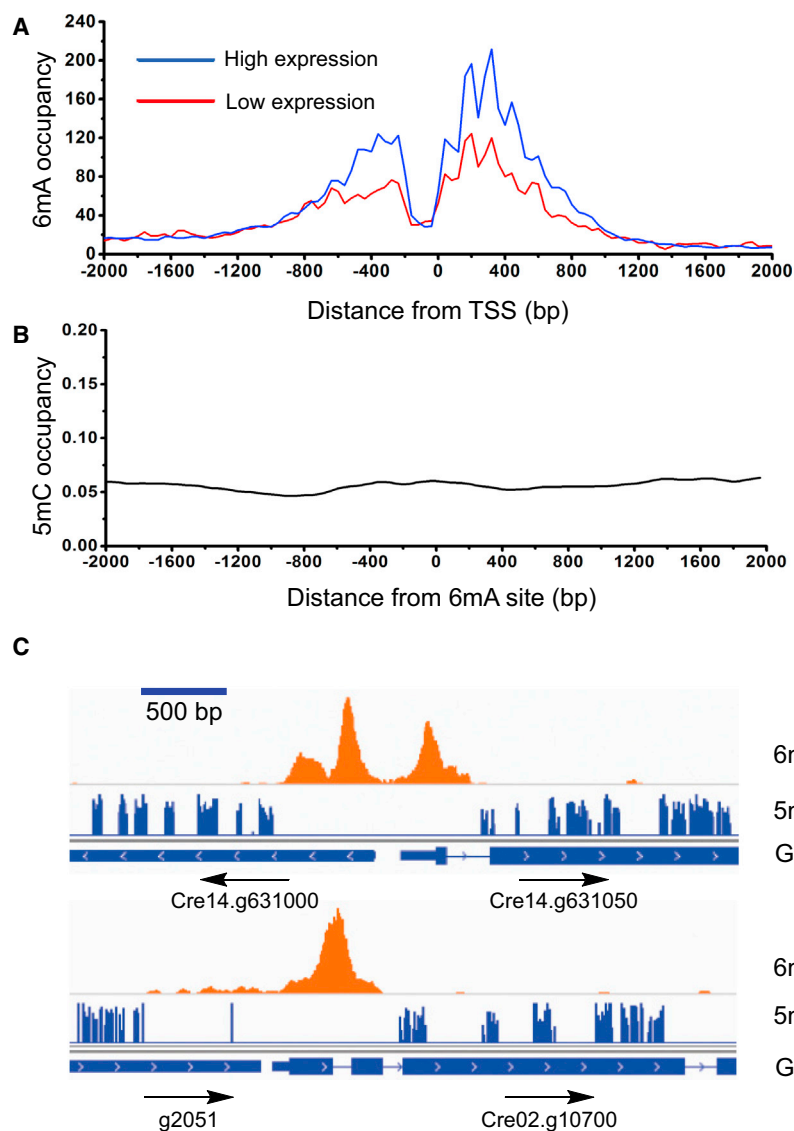


Figure 7. Correlation of 6mA with Active Genes

(A) The 6mA methylation is correlated with active genes. Two groups of genes with high (FPKM ≥ 1) and low (FPKM < 1) expression levels are plotted with its methylation level determined from 6mA-CLIP-exo-seq. 6mA occupancy represents the reads coverage that are normalized to gene counts of each category in 6mA-CLIP-exo. FPKM stands for fragments per kilobase of exon per million fragments mapped.

(B) No correlation was observed between the distributions of 5mC and 6mA. Distance between 5mC and 6mA was plotted, showing no correlation between the two.

(C) Selected examples showing that 5mC mainly appears in the gene body, whereas 6mA mainly resides near TSS region. 6mA peaks identified from 6mA-IP-seq are shown on the first line, 5mC sites identified from previous results are shown on the second line. Genome annotations are shown on the bottom line.

See also Figure S7.

DISCUSSION

6mA The 6mA and 5mC modifications are both abundant in the genome of the green algae *Chlamydomonas reinhardtii*. We showed that the total 6mA level is robustly maintained during cell proliferation. We applied 6mA-IP-seq and further developed 6mA-CLIP-exo to profile 6mA in genome wide using antibodies that specifically recognize and enrich N^6 -methylated adenine. We found that 6mA mainly resides around TSS with a bimodal distribution. The results from 6mA-CLIP-exo at higher resolution revealed that 6mA deposition occurs mainly at ApT dinucleotides within multiple sequence contexts. At least two sequence motifs,

correlated with gene expression in general (Jones, 2012); we did not observe a strong correlation between the gene expression and 5mC occupancy around TSS region (Figure S7D). This analysis indicates that 6mA and 5mC are two distinct marks in *Chlamydomonas* genome: 6mA may contribute to chromatin structures that enable initiation of gene transcription, whereas 5mC may contribute to transposon silencing, imprinting, and exon definition and affect transcription elongation (Cerutti et al., 1997).

CATG and GATC, are confirmed by a restriction enzyme digestion assay using CviAI and DpnII that are sensitive to 6mA. We then applied this restriction-enzyme-based 6mA-RE-seq strategy to *Chlamydomonas* genomic DNA and obtained genome-wide 6mA maps at single-nucleotide resolution. The identified 6mA sites within these two specific sequences account for $\sim 1/3$ of the total 6mA in genomic DNA. 6mA sites within other sequence contexts likely show similar distribution patterns (Figures 3D and S3C).

(C) Schematic models of the relationship between nucleosome distribution and 6mA in genomic DNA showing that 6mA mainly distributes in the linker DNA between two adjacent nucleosomes.

(D) Distribution profiles of 6mA and nucleosome around TSS showing that they are mostly inversely correlated.

(E) Nucleosomes exhibit a more consistent phase in relation to TSS in genes marked with 6mA than genes without 6mA.

(F) Schematic illustration of the relationship between nucleosome positioning and 6mA location in individual genes. 6mA does not reside on nucleosome-wrapped DNA.

See also Figure S6.

The results from the high-resolution maps of 6mA in two specific sequences not only validate the IP-based profiling data but also uncover a periodic pattern of 6mA. This periodicity may mark special features of transcription initiation in *Chlamydomonas* and could be related to nucleosome positioning around TSS. Indeed, we performed nucleosome footprinting coupled with high-throughput sequencing, and the results revealed a periodic pattern of nucleosome occupancy that correlates with the periodicity of 6mA distribution but is ~ 180 degrees out of phase around the TSS region. The individual 6mA sites exclusively mark the linker DNA between two adjacent nucleosomes. We propose two possible interpretations for this exclusive behavior. One possibility is that, unlike the nucleosome-wrapped DNA, the linker DNA is exposed and can thus be accessed for methylation. The other possibility is that the locations of 6mA sites contribute to the precise positioning of nucleosomes. Our results favor the latter hypothesis for the following reasons: first, we have shown that nucleosomes around TSS sites exhibit very low densities. Low-occupancy nucleosomes unlikely serve as determining factors for 6mA deposition because it occurs at almost 100% at most of these sites. On the other hand, a high density of 6mA might act to reprogram the positioning of nucleosomes around TSS regions. Second, nucleosomes are likely more dynamic than the covalent 6mA mark on DNA in the TSS regions during transcription initiation. Precedence for a role of base methylation in affecting chromatin structure exists: 5mC has been shown to contribute to nucleosome positioning in other eukaryotes (Huff and Zilberman, 2014). Additionally, 6mA may mark the TSS region for more efficient transcription initiation. Although it has been well known that the first intron is always important for transgenic gene expression in *Chlamydomonas* (Eichler-Stahlberg et al., 2009), the mechanism was unclear. We provide evidence that 6mA can reside in the first intron (examples shown in Figure 3B). The periodic distribution, its specific location on the linker DNA between two adjacent nucleosomes at TSS, and its marking of gene activation all suggest that this unique DNA mark contributes to nucleosome positioning and transcription initiation.

We have shown that 6mA shares little correlation with 5mC in the *Chlamydomonas* genome, indicating that they are controlled through different pathways and likely exhibit distinct functions. Our transcriptome analysis found an association of 6mA with gene activation; whereas, 5mC appears to negatively correlate with gene expression. Studies of 5mC have dominated notions of DNA epigenetics in eukaryotes, in particular in vertebrates, because of the critical roles played by 5mC. As shown here, 6mA can also be an important mark that could mark/affect gene activation in eukaryotes. Analogous to 5mC recognition by methyl-CpG-binding proteins, proteins that specifically recognize 6mA at TSS may exist; these proteins could interact with or be part of transcription initiation complexes that contribute to gene activation. It is also possible that 6mA may coordinate with other epigenetic factors such as histone modifications that are also enriched around the TSS region. Highly dense and narrow distributions of modifications such as H3K9 acetylation (H3K9ac) and H3K4 trimethylation (H3K4me3) near TSS have been associated with constitutive

expression of genes involved in translation in *Arabidopsis* (Ha et al., 2011). Cooperative interactions among 6mA, histone modification, and transcriptional factors could serve as a general mechanism for transcription activation in *Chlamydomonas* and possibly other eukaryotic organisms.

The *E. coli* Dam DNA methyltransferase methylates the N^6 position of adenine at GATC sites. Compared to prokaryotic 6mA modification in genomic DNA, the 6mA methylation in the *Chlamydomonas* genome exists in a more complex manner with multiple potential sequences mainly centered on ApT, resembling eukaryotic 5mC methylation of CpG. The methyltransferases that are involved in establishing or maintaining the patterns of 6mA sites remain to be determined (Iyer et al., 2011). It should be noted that Greer et al. (2015) (this issue of *Cell*) have recently discovered two enzymes, MAD-1 and DMT-1, which can install or remove 6mA in the genome of *Caenorhabditis elegans*, respectively.

In summary, our study has demonstrated that 6mA is an abundant DNA modification in the *Chlamydomonas* genome. It is enriched specifically around TSS and preferentially marks actively transcribed genes. A periodic distribution pattern with depletion at the TSS coupled with an almost exclusive marking of the linker DNA between adjacent nucleosomes indicates a process of controlled deposition, as well as functional roles in nucleosome positioning and transcriptional initiation. Although 5mC is well known to mark gene repression at promoter and enhancer sites in vertebrates, we show in this work that a different DNA base modification, 6mA, flanks TSS and marks actively transcribed genes. The ribose version of 6mA modification (with 2'-OH) exists as the most abundant internal mRNA modification in almost all eukaryotes. It has recently been shown to be reversible and plays important regulatory functions (Fu et al., 2014). We suspect that 6mA could be widely present in eukaryotic genomes as well; in certain species, 6mA may carry important roles in regulating gene expression; in other organisms, 6mA may play complementary roles to 5mC at different stages of development.

EXPERIMENTAL PROCEDURES

6mA-IP-Seq

Isolated genomic DNA was diluted to 100–200 ng/ μ l using TE buffer and sonicated in 130 μ l scale to 200–400 bp using a Covaris Focused-ultrasonicator. End repair, 3'-adenylation, and adaptor ligation were performed. The ligated and purified DNA was denatured at 95°C and chilled on ice. A portion of 10 μ l DNA was saved as input. The rest of the DNA was combined with 3 μ g of anti-6mA antibody (Synaptic Systems) in 500 μ l of 1 \times IP buffer and incubated at 4°C for 6 hr. At the same time, 40 μ l of Protein A magnetic beads was washed twice in 0.5 ml of 1 \times IP buffer and pre-blocked in 0.5 ml of 1 \times IP buffer containing 20 μ g/ μ l of Bovine Serum Albumin at 4°C for 6 hr. The Protein A beads were washed twice with 0.5 ml of 1 \times IP buffer, added to the DNA-6mA antibody mixture, and incubated at 4°C with gentle rotation overnight. The beads were then washed four times with 0.5 ml of 1 \times IP buffer. Methylated DNA was eluted twice by 100 μ l of elution buffer containing 6mA monophosphate at 4°C for 1 hr. The two elution solutions were combined, to which 20 μ l of NaOAc (3 M, pH 5.3), 500 μ l of EtOH, and 0.5 μ l of glycogen (20 ng/ μ l) were added. The solution was frozen at -80°C overnight and centrifuged at 14,000 $\times g$ for 20 min at 4°C. The precipitated DNA was dissolved in 7 μ l of ddH₂O, PCR amplified for 15 to 18 cycles, purified by Ampure beads, and suspended in 16 μ l of re-suspension buffer to yield the sequencing library.

6mA-CLIP-Exo

Genomic DNA was sonicated to around 200 bp and immunoprecipitated by using anti-6mA antibody. The antibody-DNA complex was then covalently crosslinked using UV 254 nm irradiation, followed by a procedure similar to ChIP-exo (Rhee and Pugh, 2012). The library was constructed with Illumina-compatible adapters and primers and applied to Illumina HiSeq 2000 sequencer with single-end reads. The raw data were aligned by bowtie, and the peaks were called by MACE (model-based analysis of ChIP-exo). See the Extended Experimental Procedures for detailed procedures.

6mA-RE-Seq

Restriction enzyme digestion was performed by treating 1 μ g of gDNA with 5 μ l of CviAI or DpnII restriction enzyme (5 U/ μ l) at 25°C or 37°C overnight. The digested and non-digested DNA (200 ng each) were fragmented into ~100 bp by sonication, and sequencing libraries were constructed according to Illumina TruSeq DNA sample preparation procedures.

Detection of 6mA Peaks from 6mA-IP-Seq and 6mA-RE-Seq

Reads were mapped to the *Chlamydomonas* genome (JGI) Version 9.1, with parameters and scripts as described in the Extended Experimental Procedures.

ACCESSION NUMBERS

Sequencing data have been deposited into the Gene Expression Omnibus (GEO) under the accession number GSE62690.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, seven figures, and one table and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.04.010>.

AUTHOR CONTRIBUTIONS

Y.F., G.-Z.L., and C.H. conceived the project. Y.F., G.-Z.L., and K.C. designed and performed experiments with help from, X.D., M.Y., Z.H., J.L., X.L., X.W., and Q.J. G.-Z.L. conducted bioinformatics analysis. D.H. helped on data analysis. L.C.D. helped on high-throughput sequencing, and L.M. helped with experiment design. Y.F., G.-Z.L., and C.H. wrote the manuscript with input from K.C. and L.M.

ACKNOWLEDGMENTS

We thank Dr. Q. Jin for help with HPLC-QQQ-MS/MS and Dr. Pieter W. Faber for help with high-throughput sequencing. This work is partially supported by National Institutes of Health (R01 HG006827 to C.H.), and C.H. is a Howard Hughes Medical Institute investigator. The Mass Spectrometry Facility of the University of Chicago is funded by National Science Foundation (CHE-1048528). S.F. Reichard, MA edited the manuscript.

Received: October 26, 2014

Revised: February 16, 2015

Accepted: March 27, 2015

Published: April 30, 2015

REFERENCES

Babinger, P., Kobl, I., Mages, W., and Schmitt, R. (2001). A link between DNA methylation and epigenetic silencing in transgenic *Volvox carteri*. *Nucleic Acids Res.* 29, 1261–1271.

Bird, A. (2007). Perceptions of epigenetics. *Nature* 447, 396–398.

Bisova, K., Krylov, D.M., and Umen, J.G. (2005). Genome-wide annotation and expression profiling of cell cycle regulatory genes in *Chlamydomonas reinhardtii*. *Plant Physiol.* 137, 475–491.

Cerutti, H., Johnson, A.M., Gillham, N.W., and Boynton, J.E. (1997). Epigenetic silencing of a foreign gene in nuclear transformants of *Chlamydomonas*. *Plant Cell* 9, 925–945.

Chen, K., Lu, Z., Wang, X., Fu, Y., Luo, G.-Z., Liu, N., Han, D., Dominissini, D., Dai, Q., Pan, T., and He, C. (2015). High-resolution N(6)-methyladenosine (m⁶A) map using photo-crosslinking-assisted m⁶A sequencing. *Angew. Chem. Int. Ed. Engl.* 54, 1587–1590.

Chodavarapu, R.K., Feng, S., Bernatavichute, Y.V., Chen, P.-Y., Stroud, H., Yu, Y., Hetzel, J.A., Kuo, F., Kim, J., Cokus, S.J., et al. (2010). Relationship between nucleosome positioning and DNA methylation. *Nature* 466, 388–392.

Collier, J., McAdams, H.H., and Shapiro, L. (2007). A DNA methylation ratchet governs progression through a bacterial cell cycle. *Proc. Natl. Acad. Sci. USA* 104, 17111–17116.

Cummings, D.J., Tait, A., and Goddard, J.M. (1974). Methylated bases in DNA from *Paramecium aurelia*. *Biochim. Biophys. Acta* 374, 1–11.

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., et al. (2012). Topology of the human and mouse m⁶A RNA methylomes revealed by m⁶A-seq. *Nature* 485, 201–206.

Ehrlich, M., and Zhang, X.-Y. (1990). Naturally Occurring Modified Nucleosides in DNA. In *Journal of Chromatography Library, Chapter 10*, Charles W.G. and Kenneth C.T.K., eds. (Elsevier), pp. B327–B362.

Eichler-Stahlberg, A., Weisheit, W., Ruecker, O., and Heitzer, M. (2009). Strategies to facilitate transgene expression in *Chlamydomonas reinhardtii*. *Planta* 229, 873–883.

Engel, J.D., and von Hippel, P.H. (1978). Effects of methylation on the stability of nucleic acid conformations. Studies at the polymer level. *J. Biol. Chem.* 253, 927–934.

Fang, G., Munera, D., Friedman, D.I., Mandlik, A., Chao, M.C., Banerjee, O., Feng, Z., Losic, B., Mahajan, M.C., Jabado, O.J., et al. (2012). Genome-wide mapping of methylated adenine residues in pathogenic *Escherichia coli* using single-molecule real-time sequencing. *Nat. Biotechnol.* 30, 1232–1239.

Feng, S., Cokus, S.J., Zhang, X., Chen, P.Y., Bostick, M., Goll, M.G., Hetzel, J., Jain, J., Strauss, S.H., Halpern, M.E., et al. (2010). Conservation and divergence of methylation patterning in plants and animals. *Proc. Natl. Acad. Sci. USA* 107, 8689–8694.

Fu, Y., Dominissini, D., Rechavi, G., and He, C. (2014). Gene expression regulation mediated through reversible m⁶A RNA methylation. *Nat. Rev. Genet.* 15, 293–306.

Greer, E.L., Blanco, M.A., Gu, L., Sendinc, E., Liu, J., Aristizábal-Corralles, D., Hsu, C.-H., Aravind, L., He, C., and Shi, Y. (2015). DNA methylation on N⁶-adenine in *C. elegans*. *Cell* 161, this issue, 868–878.

Ha, M., Ng, D.W., Li, W.H., and Chen, Z.J. (2011). Coordinated histone modifications are associated with gene expression variation within and between species. *Genome Res.* 21, 590–598.

Hattman, S., Kenny, C., Berger, L., and Pratt, K. (1978). Comparative study of DNA methylation in three unicellular eucaryotes. *J. Bacteriol.* 135, 1156–1157.

Huff, J.T., and Zilberman, D. (2014). Dnmt1-independent CG methylation contributes to nucleosome positioning in diverse eukaryotes. *Cell* 156, 1286–1297.

Iyer, L.M., Abhimani, S., and Aravind, L. (2011). Natural history of eukaryotic DNA methylation systems. *Prog. Mol. Biol. Transl. Sci.* 101, 25–104.

Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., Yi, C., Lindahl, T., Pan, T., Yang, Y.G., and He, C. (2011). N6-methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. *Nat. Chem. Biol.* 7, 885–887.

Jones, P.A. (2012). Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* 13, 484–492.

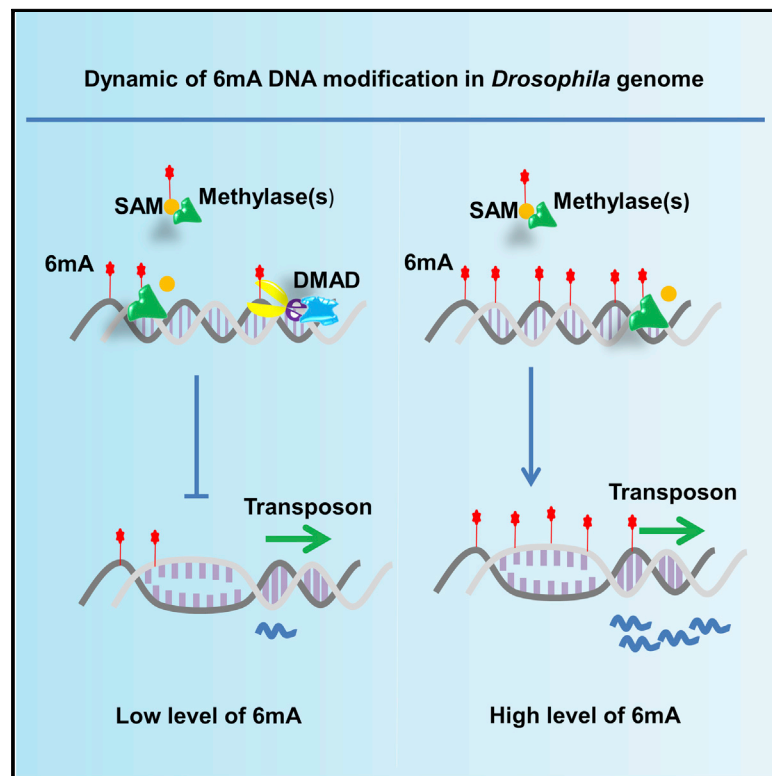
Law, J.A., and Jacobsen, S.E. (2010). Establishing, maintaining and modifying DNA methylation patterns in plants and animals. *Nat. Rev. Genet.* 11, 204–220.

Lodha, M., and Schroda, M. (2005). Analysis of chromatin structure in the control regions of the *chlamydomonas* HSP70A and RBCS2 genes. *Plant Mol. Biol.* 59, 501–513.

- Low, D.A., Weyand, N.J., and Mahan, M.J. (2001). Roles of DNA adenine methylation in regulating bacterial gene expression and virulence. *Infect. Immun.* 69, 7197–7204.
- Lu, M., Campbell, J.L., Boye, E., and Kleckner, N. (1994). SeqA: a negative modulator of replication initiation in *E. coli*. *Cell* 77, 413–426.
- Merchant, S.S., Prochnik, S.E., Vallon, O., Harris, E.H., Karpowicz, S.J., Witman, G.B., Terry, A., Salamov, A., Fritz-Laylin, L.K., Maréchal-Drouard, L., et al. (2007). The *Chlamydomonas* genome reveals the evolution of key animal and plant functions. *Science* 318, 245–250.
- Messer, W., and Noyer-Weidner, M. (1988). Timing and targeting: the biological functions of Dam methylation in *E. coli*. *Cell* 54, 735–737.
- Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E., and Jaffrey, S.R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* 149, 1635–1646.
- Murray, I.A., Clark, T.A., Morgan, R.D., Boitano, M., Anton, B.P., Luong, K., Fomenkov, A., Turner, S.W., Korlach, J., and Roberts, R.J. (2012). The methylomes of six bacteria. *Nucleic Acids Res.* 40, 11450–11462.
- Radakovits, R., Jinkerson, R.E., Darzins, A., and Posewitz, M.C. (2010). Genetic engineering of algae for enhanced biofuel production. *Eukaryot. Cell* 9, 486–501.
- Rae, P.M., and Spear, B.B. (1978). Macronuclear DNA of the hypotrichous ciliate *Oxytricha fallax*. *Proc. Natl. Acad. Sci. USA* 75, 4992–4996.
- Ratel, D., Ravanat, J.L., Berger, F., and Wion, D. (2006). N6-methyladenine: the other methylated base of DNA. *BioEssays* 28, 309–315.
- Rhee, H.S., and Pugh, B.F. (2012). ChIP-exo method for identifying genomic location of DNA-binding proteins with near-single-nucleotide accuracy. *Curr. Protoc. Mol. Biol.*, Unit 21.24.
- Rogers, J.C., and Rogers, S.W. (1995). Comparison of the effects of N6-methyldeoxyadenosine and N5-methyldeoxycytosine on transcription from nuclear gene promoters in barley. *Plant J.* 7, 221–233.
- Sasaki, H., and Matsui, Y. (2008). Epigenetic events in mammalian germ-cell development: reprogramming and beyond. *Nat. Rev. Genet.* 9, 129–140.
- Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* 14, 204–220.
- Sternberg, N. (1985). Evidence that adenine methylation influences DNA-protein interactions in *Escherichia coli*. *J. Bacteriol.* 164, 490–493.
- Vovis, G.F., and Lacks, S. (1977). Complementary action of restriction enzymes endo R-DpnI and Endo R-DpnII on bacteriophage f1 DNA. *J. Mol. Biol.* 115, 525–538.
- Weber, M., Davies, J.J., Wittig, D., Oakeley, E.J., Haase, M., Lam, W.L., and Schübeler, D. (2005). Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. *Nat. Genet.* 37, 853–862.
- Wion, D., and Casadesús, J. (2006). N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. *Nat. Rev. Microbiol.* 4, 183–192.
- Zhang, Y., Nelson, M., Nietfeldt, J.W., Burbank, D.E., and Van Etten, J.L. (1992). Characterization of *Chlorella virus* PBCV-1 CviAll restriction and modification system. *Nucleic Acids Res.* 20, 5351–5356.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoutte, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., and Liu, X.S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 9, R137.

N⁶-Methyladenine DNA Modification in *Drosophila*

Graphical Abstract



Authors

Guoqiang Zhang, Hua Huang, ...,
Hailin Wang, Dahua Chen

Correspondence

hlwang@rcees.ac.cn (H.W.),
chendh@ioz.ac.cn (D.C.)

In Brief

DNA N⁶-methyladenine (6mA) modification is present in the *Drosophila* genome. The *Drosophila* Tet-homolog-mediated 6mA demethylation correlates with transposon expression, suggesting a potential role of 6mA in regulating gene expression.

Highlights

- 6mA modification occurs in fly embryo DNA and is regulated by DMAD
- DMAD promotes differentiation of early germ cells in fly ovary
- DMAD catalyzes demethylation of 6mA both in vivo and in vitro
- DMAD-mediated 6mA demethylation is correlated with transposon expression

Accession Numbers

SRP055483



N⁶-Methyladenine DNA Modification in *Drosophila*

Guoqiang Zhang,^{1,5} Hua Huang,^{2,5} Di Liu,^{2,5} Ying Cheng,¹ Xiaoling Liu,² Wenxin Zhang,¹ Ruichuan Yin,² Dapeng Zhang,² Peng Zhang,³ Jianzhao Liu,⁴ Chaoyi Li,¹ Baodong Liu,² Yuewan Luo,¹ Yuanxiang Zhu,¹ Ning Zhang,² Shunmin He,³ Chuan He,⁴ Hailin Wang,^{2,*} and Dahua Chen^{1,*}

¹State Key Laboratory of Reproductive Biology, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China

²State Key Laboratory of Environmental Chemistry and Ecotoxicology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China

³Key Laboratory of the Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, Beijing 100101, China

⁴Department of Chemistry and Institute for Biophysical Dynamics, Howard Hughes Medical Institute, The University of Chicago, Chicago, IL 60637, USA

⁵Co-first author

*Correspondence: hlwang@rcees.ac.cn (H.W.), chendh@ioz.ac.cn (D.C.)

<http://dx.doi.org/10.1016/j.cell.2015.04.018>

SUMMARY

DNA N⁶-methyladenine (6mA) modification is commonly found in microbial genomes and plays important functions in regulating numerous biological processes in bacteria. However, whether 6mA occurs and what its potential roles are in higher-eukaryote cells remain unknown. Here, we show that 6mA is present in *Drosophila* genome and that the 6mA modification is dynamic and is regulated by the *Drosophila* Tet homolog, DNA 6mA demethylase (DMAD), during embryogenesis. Importantly, our biochemical assays demonstrate that DMAD directly catalyzes 6mA demethylation in vitro. Further genetic and sequencing analyses reveal that DMAD is essential for development and that DMAD removes 6mA primarily from transposon regions, which correlates with transposon suppression in *Drosophila* ovary. Collectively, we uncover a DNA modification in *Drosophila* and describe a potential role of the DMAD-6mA regulatory axis in controlling development in higher eukaryotes.

INTRODUCTION

DNA methylation, an epigenetic mechanism, does not change DNA sequence but instead suppresses the transcription factor-DNA association, thereby regulating gene expression and a variety of cellular processes (Feng et al., 2010; Smith and Meissner, 2013). Several methylated bases, including 5-methylcytosine (5mC), N⁶-methyladenine (6mA), and N⁴-methylcytosine (4mC), have been found in genomic DNA from diverse species (Cheng, 1995; Ratel et al., 2006; Wion and Casadesús, 2006). These methylated bases have been shown to be products of post-replicative DNA modification generated by specific DNA methylases (Wion and Casadesús, 2006). The prevailing view is that, unlike 5mC, 6mA and 4mC function only in bacteria, protists, and other lower eukaryotes (Cheng, 1995; Wion and Casadesús, 2006). Among these DNA modifications, 6mA plays an important role in controlling a number of biological functions

in bacteria, such as DNA replication and repair, gene expression, and host-pathogen interactions (Reisenauer et al., 1999; Wion and Casadesús, 2006), and is essential for viability of some bacterial strains (Julio et al., 2001; Stephens et al., 1996; Wright et al., 1997). In contrast, 5mC is thought to be the predominant type, if not the only type, of methylated base in mammals (Smith and Meissner, 2013).

Recent studies have suggested that methylation/demethylation at the C-5 position of cytosine in mammals is a dynamic and reversible process controlled by several mechanisms, including passive and active demethylation (Bhutani et al., 2011; Wu and Zhang, 2014). While passive demethylation is attributed to successive cell divisions that cause a progressive loss of 5mC on a genome scale, active demethylation is achieved by ten-eleven translocation (Tet)-mediated oxidation to 5-hydroxymethylcytosine (5hmC) (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). It has been shown that 5mC can be oxidized by the Tet enzymes in an iterative manner to 5hmC, 5-formylcytosine (5fC), and 5-carboxylcytosine (5caC), and both 5fC and 5caC can be further replaced to unmodified cytosine by excision repair pathway (He et al., 2011; Ito et al., 2011; Maiti and Drohat, 2011).

Given the important roles of 6mA modification in bacteria, we explore whether 6mA plays a role in eukaryotes. However, previous studies have suggested that the 6mA base is present at extremely low levels in genomic DNA of higher eukaryotes (Ratel et al., 2006). We speculate that, if 6mA plays a role, the potential installation of this modification by methyltransferases could be reversed by a demethylase-mediated demethylation process. Since extremely low levels of 6mA are present in higher eukaryotes, we reason that 6mA demethylases might play predominant roles in controlling the dynamics of 6mA DNA modification in higher eukaryotes. Thus, knockout of yet-to-be identified demethylases could lead to accumulation of 6mA and allow its functional investigations. In this study, we identify the *Drosophila* Tet homolog as the DNA demethylase that is responsible for 6mA demethylation in *Drosophila*, and we name it DNA 6mA demethylase (DMAD).

In *Drosophila*, 5mC modification exists at an extremely low level (Lyko and Maleszka, 2011), and the *Drosophila* Dnmt2-dependent methylome lacks defined DNA 5mC patterns (Radatz et al., 2013). Thus, whether the *Drosophila* genome has a

functional DNA modification remains elusive. In this study, we show that DNA modification 6mA is present in the *Drosophila* genome at a considerable level and that the demethylation of 6mA is tightly regulated by DMAD during embryogenesis and tissue homeostasis. We also demonstrate that DMAD is likely a 6mA demethylase since it directly catalyzes 6mA demethylation in vitro. Further genetic and sequencing analyses reveal that DMAD determines 6mA distribution in the *Drosophila* genome and is essential for development.

RESULTS

Characterization of 6mA Modification in *Drosophila* Genomic DNA

Previous studies suggested that 5mC modification in *Drosophila* DNA occurs at very low levels, and the *Drosophila* Dnmt2-dependent methylome lacks defined DNA 5mC methylation patterns (Lyko et al., 2000; Raddatz et al., 2013). To further explore this issue, we employed ultra-high-performance liquid chromatography-triple quadrupole mass spectrometry, coupled with multiple-reaction monitoring (UHPLC-MRM-MS/MS) analysis, an extremely sensitive assay for detecting base modification (Yin et al., 2013), to measure the abundance of oxidized 5mC derivatives, 5hmC, 5fC, and 5caC in multiple tissues. The UHPLC-MRM-MS/MS assays showed that, although 5hmC was detected in *Drosophila* DNA at extremely low levels and fewer than 100 of the cytosine bases per genome were modified to be 5hmC (Figures S1A–S1C), 5fC and 5caC were not detectable in *Drosophila* DNA. These observations prompted us to explore whether DNA methylation could occur at other bases. We turned our attention to explore the possible existence of adenine methylation in *Drosophila* DNA.

We used an antibody that is specifically against the 6mA base in DNA (Figure S1D) and performed dot blot experiments to detect the 6mA signal in *Drosophila* DNA samples isolated from various adult tissues and from embryos at various stages. As shown in Figure 1A, while relatively weak signals of 6mA were detected in DNA from adult tissues and late-stage embryos, a very strong 6mA signal was found to be present in embryos at the very early stage, suggesting the existence of 6mA in *Drosophila* DNA and that the status of 6mA modification might be dynamic during embryogenesis.

We next sought to quantify 6mA in *Drosophila* DNA using the UHPLC-MRM-MS/MS method (Figure S1E) and first focused on measuring the 6mA abundance in DNA at the embryonic stages. As shown in Figures 1B and 1C, abundance of the 6mA base appeared to display a peak ($\sim 0.07\%$, 6mA/dA) at the ~ 0.75 hr stage but was dramatically reduced to a very low level ($\sim 0.001\%$, 6mA/dA) at the 4–16 hr stages, confirming that 6mA is dynamic in *Drosophila* DNA during embryonic development. Additionally, we also quantified the abundance of 6mA in adult tissues (e.g., brain and ovary) and found that it exhibited similar low levels to those found in the late-stage embryonic genome (Figures 1D and 1E). To confirm that the signal indeed reveals 6mA modification in *Drosophila*, we collected the peak fraction containing 6mA (Figure S1F) and performed a further high-resolution mass spectrometry analysis. As shown in Figures 1F–1H, we observed an accurate mass/charge ratio of

266.1250 au (M+H), which matched the theoretic monoisotopic mass of 6mA (266.1248 au) with a deviation of 1.02 ppm. Notably, the isolated compound displayed the same fragment pattern (20 fragments) as the standard 6mA. Collectively, our findings support that 6mA is present in fly DNA and is highly dynamic during early embryogenesis.

Drosophila Embryos Possess DNA 6mA Demethylation Activity

The observation of a dramatic reduction in 6mA levels in the *Drosophila* genome from the very early to the late stages of embryonic development prompted us to ask the intriguing question of whether active 6mA demethylation occurs during *Drosophila* embryogenesis. To explore this issue, we established an in vitro DNA 6mA demethylation assay. In this assay, we employed the AlkB, a known 6mA demethylase from bacteria (Li et al., 2012), as a positive control enzyme (Figure 2A). As shown in Figure 2B, contrary to the control reaction with adding the GFP protein, the methylated DNA substrates were significantly oxidized in the presence of AlkB in a dose-dependent manner. We then used this established system to determine whether the embryonic nuclear extracts have enzymatic activity for 6mA demethylation. As shown in Figure 2C, addition of serially diluted nuclear extracts in the enzymatic reaction catalyzed 6mA demethylation in a dose-dependent manner. By contrast, no or a low background signal of 6mA demethylation was measured in the control reactions in which GFP was added. Of note, we found that no or only a low level of background demethylation signal was detected when we added the same number of boiled nuclear extracts in a parallel control reaction (data not shown), suggesting that a potential small amount of DNA from nuclear extracts did not interfere with the signal that we collected from in vitro reactions.

Interestingly, we detected an increase in 6mA demethylation activity from nuclear extracts during embryonic development. As shown in enzymatic assays, the 6mA demethylation activity of nuclear extracts was relative low at the very early stage but gradually increased and reached a peak at the 6 hr stage (Figure 2D). This result demonstrated that 6mA demethylation activity and abundance of 6mA in embryonic DNA are mutually complementary with each other during embryonic development (see Figures 1A, 1C, and 2D). Thus, our findings not only support that DNA 6mA modification is a dynamic process during early *Drosophila* embryonic development, but also raise a possibility that 6mA demethylation is regulated by a specific DNA dioxygenase.

DMAD Is Involved in Regulating DNA 6mA Demethylation

We next aimed to search for the specific enzyme responsible for 6mA demethylation. Previous studies have shown that Tet proteins in mammals play important roles in DNA demethylation through converting 5mC to 5hmC (Kriaucionis and Heintz, 2009; Tahiliani et al., 2009). The *Drosophila* genome contains a gene, CG2083, which encodes a putative dioxygenase protein. Sequence alignment and domain structure analysis suggested that this protein contains highly conserved domains, including a CXXC zinc finger (645 aa–684 aa), a Cys-rich domain (1695 aa–1867 aa), and a DSBH domain (1888 aa–2918 aa),

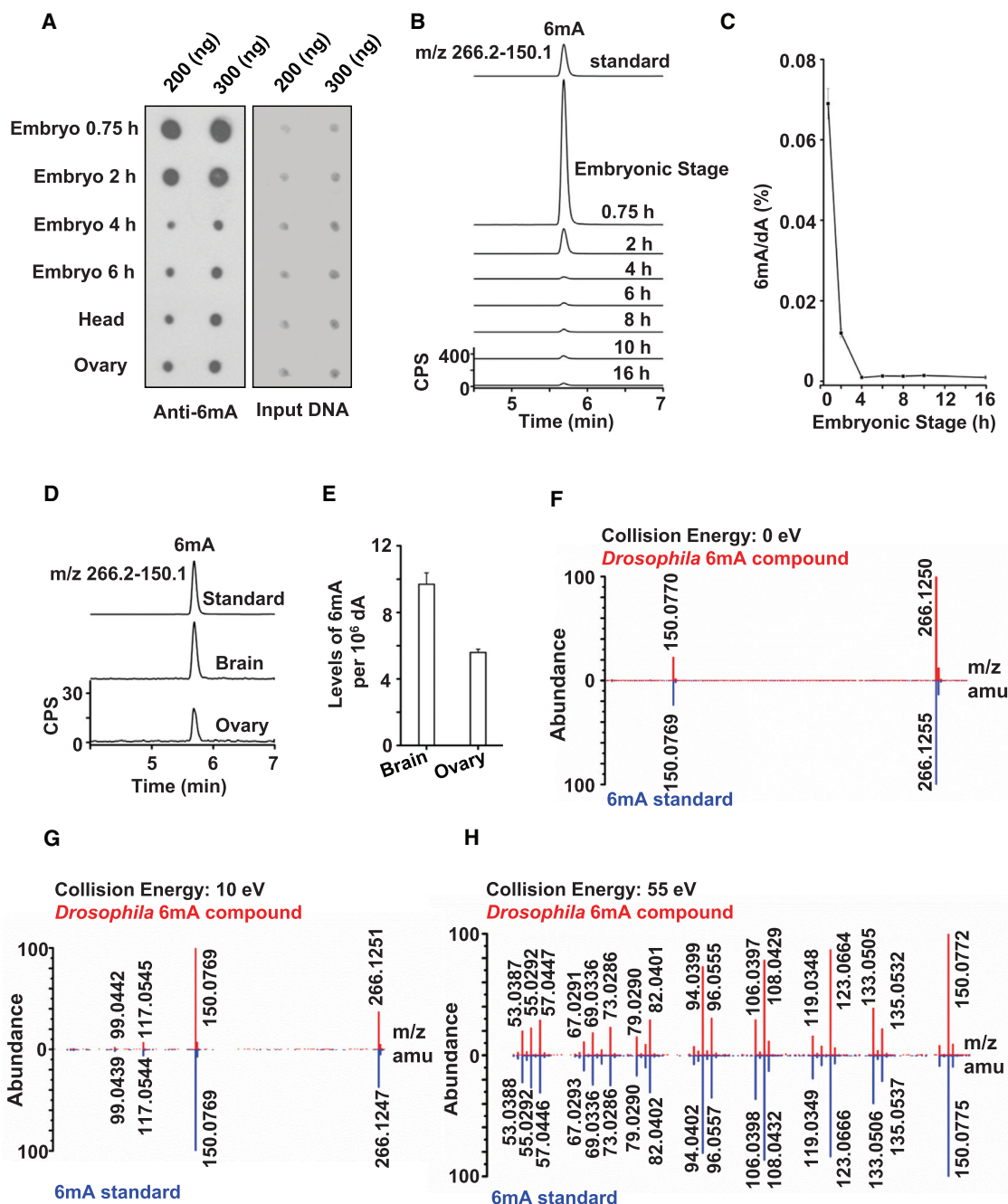


Figure 1. Characterization and Quantification of 6mA in *Drosophila* DNA

(A) Genomic DNAs from embryos at various stages and adult tissues as indicated were subjected to dot blot assays using a specific anti-6mA antibody (left). Methylene blue hydrate staining was performed to determine the signal of input DNA (right). (B–E) UHPLC-MRM-MS/MS chromatograms (B and D) and quantification (C and E) of 6mA in genomic DNA of embryos (B and C) and adult tissues (D and E). (F–H) Control compound and the isolated 6mA compound from fly genomic DNA from 0.5–1.5 hr embryos were subjected to further high-resolution mass-spectrometry analysis. The collision energy was set at 0 eV (F), 10 eV (G), and 55 eV (H).

The experiments were carried out by triplicates, and the standard deviations were calculated by Excel. See also Figure S1.

which are also present in mammalian Tet proteins (Figure 2E). It is worthwhile to note that the bacterial AlkB protein also contains a DSBH domain (Figure 2E). Based on the biochemical function of the CG2083-encoding protein that we characterized below,

we thereafter called it *Drosophila* DNA 6mA demethylase and abbreviated it as DMAD. Given that DMAD looks more like mammalian Tet proteins, we first performed an in vitro enzymatic assay and found that the catalytic domain of DMAD (DMAD-CD),

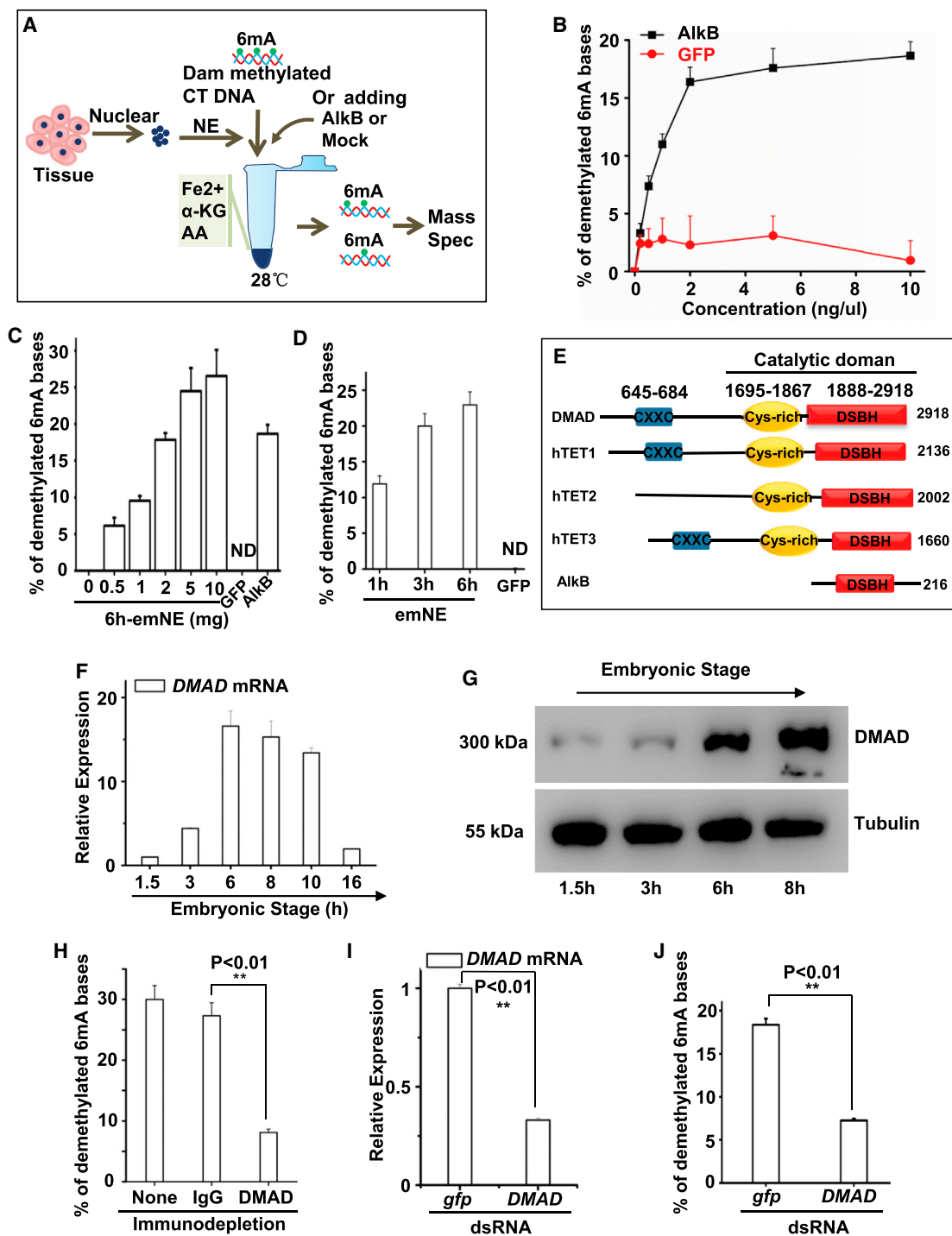


Figure 2. Demethylation of DNA 6mA Modification by *Drosophila* Embryo Nuclear Extracts

(A) Schematic representation of in vitro 6mA demethylation. AlkB- and mock-treated nuclear extracts were used in the enzymatic reaction to catalyze the demethylation of the 6mA in methylated CT DNA, and the products were then subjected to mass spectrometry.

(B) *E. coli* AlkB, but not mock protein (GFP), catalyzes the demethylation of the 6mA in a dose-dependent fashion.

(C) Nuclear extracts from 6 hr embryos (from 0.5, 1, 2, 5, 10 mg embryos as indicated) were tested in the demethylation reaction. In this experiment, AlkB and GFP were used as positive or negative controls, respectively.

(D) Nuclear extracts (from 2 mg embryos) from various embryonic stages possess enzyme activity for 6mA demethylation.

(E) Schematic diagram showing that DMAD contains three conserved domains—the CXXC zinc finger, the Cys-rich domain, and the DSBH domain—that are also present in mammalian Tet proteins. The bacterium AlkB contains DSBH, which possesses enzyme activity for 6mA demethylation.

(legend continued on next page)

but not DMAD-CD^{mut}, could convert 5mC to 5hmC in vitro (Figures S2A–S2D). However, the in vitro activity of DMAD-CD that mediates oxidation of 5mC was about 30-fold lower than that of mouse Tet1-CD (data not shown). Nevertheless, only a few cytosine bases (less than 100) as mentioned above could be detected as the hydroxymethylated form, 5hmC, per genome (Figures S1B and S1C). We speculated that DMAD might play a role in catalyzing other forms of DNA modification in fly, for example, 6mA.

We found that the DMAD is weakly expressed during early embryonic stages but is highly expressed during the later embryonic stages (Figures 2F and 2G). Thus, the DMAD expression has a complementary expression pattern as 6mA during embryogenesis. Because nuclear extracts from the late-stage embryos exhibited considerable 6mA demethylation activity and high levels of DMAD expression, we asked whether DMAD is involved in regulating the 6mA demethylation. To do this, we used a specific anti-DMAD antibody and then performed antibody-depletion experiments. Nuclear extracts with depleted DMAD were then used in in vitro 6mA DNA demethylation assays. As shown in Figure 2H, depletion by the anti-DMAD antibody, but not IgG, significantly blocked the demethylation activity of the nuclear extracts from the late-stage embryos, arguing that the DMAD is involved in regulating 6mA demethylation.

To obtain further evidence to support our argument, we next employed the double-strand RNA (dsRNA) knockdown method and further evaluated the specificity of DMAD's role in regulating 6mA demethylation. As shown in Figure 2I, injection of dsRNA against the *DMAD* mRNA in embryos significantly reduced the *DMAD* expression. As shown in Figure 2J, nuclear extracts from embryos treated with *DMAD* dsRNA exhibited much less 6mA demethylation activity than extracts from control embryos, further confirming the important role of DMAD in demethylating 6mA. In line with this, we observed that knockdown of DMAD increased the levels of 6mA in late-stage (15 hr) embryos (Figure S2E). In addition, we found that injection of *DMAD* dsRNA at different developmental time points caused significant lethality at the late embryonic stage when compared with control dsRNA injection (Figure S2F), suggesting that DMAD possibly contributes to embryonic development.

DMAD Is Required for *Drosophila* Development

To investigate the biological role of DMAD and its relevance to fly DNA 6mA demethylation in vivo, we sought to generate the *DMAD* mutant flies by employing the CRISPR/Cas system. According to the method described previously (Cong et al., 2013; Mali et al., 2013), we designed two sgRNAs containing non-overlapping sequences targeting the *DMAD* gene and generated two alleles, *DMAD*¹ and *DMAD*², with an independent genetic background (Figure 3A and Extended Experimental Procedures). As shown in a western blot assay, DMAD expression

was completely abolished in the *DMAD*¹ and *DMAD*² mutant allelic combination (Figure 3B), revealing that these two *DMAD* mutants are null alleles. To determine the biological role of DMAD, we performed a genetic complementation test and found that, while most of *trans*-heterozygous mutant animals were lethal at the pupa stage, a small population of mutant animals were able to pass through the pupa stage but died within 3 days post-eclosion (Figure 3C).

We next determined the role of DMAD in demethylating 6mA in vivo. We prepared genomic DNA from both wild-type and *DMAD*^{1/2} mutant flies and measured the abundance of the 6mA base. As shown in Figure 3D, loss of *DMAD* led to a significant increase in the overall 6mA abundance in genomic DNA. Of note, we found no difference in the abundance of 5mC and 5hmC between the wild-type and *DMAD* mutant flies (Figures 3E, S3A, and S3B), strongly arguing that the *Drosophila* DMAD has no apparent in vivo role in regulating the conversion of 5mC to 5hmC. Moreover, we found that, while N³-methylcytosine (3mC) and O⁶-methylguanine (m6G) were not detectable, N¹-methyladenine (1mA) (below 0.6 adduct per million dA) and N³-methyladenosine (3mA) (about 2 adducts per million dA) were present at low levels in both wild-type and *DMAD* mutant flies, and no difference in relative abundance of 1mA and 3mA bases was detected between wild-type and *DMAD* mutant flies (Figure S3C; see Discussion). Additionally, we failed to detect any apparent difference in levels of m6A abundance in RNA between wild-type and *DMAD* mutant flies (Figure S3D). These results together suggest that the DMAD specifically suppresses the in vivo modification of 6mA, rather than 5mC and other methylated DNA bases tested in this study, and m6A in RNA.

We then sought to determine the functional requirements of the conserved domains (Figure S3F) in the DMAD protein by generating specific domain-deletion alleles. To do so, we designed two additional sgRNAs and attempted to use the Cas9/sgRNA technique to locally produce truncated proteins of the DMAD (Figure 3F). According to the experimental design, we successfully obtained two new *DMAD* alleles, *DMAD*^{del-CXXC} and *DMAD*^{del-CD}. These two alleles encode putative truncated proteins, in which the CXXC domain and the catalytic domain were deleted in DMAD, respectively (Figure 3F and Figure S3E).

Our genetic experiments showed that *DMAD*^{del-CXXC} homozygous mutant animals are viable and fertile and that the *DMAD*^{del-CXXC} allele is able to complement both *DMAD*¹ and *DMAD*² alleles (Figure 3C). In addition, UHPLC-MRM-MS/MS assays showed that the CXXC domain deletion did not cause significant change in 6mA abundance in DNA between *DMAD*^{del-CXXC} homozygous and wild-type flies (Figure 3G). Of note, in our western blot assays, we found that wild-type flies also expressed a similar size protein as present in *DMAD*^{del-CXXC} homologue (Figure S3E). Taken together, these results suggested that the CXXC domain is dispensable for the

(F) Expression levels of *DMAD* at different embryonic stages as measured by qRT-PCR.

(G) Expression levels of DMAD protein at different embryonic stages measured by western blot assays.

(H) Nuclear extracts treated with anti-DMAD or IgG or without treatment were used in in vitro 6mA demethylation assays.

(I) Relative expression levels of *DMAD* in embryos treated with dsRNA against *DMAD* or *gfp* were measured by qRT-PCR.

(J) Nuclear extracts from embryos treated with dsRNA against *DMAD* or *gfp* were used in in vitro 6mA DNA demethylation assays.

The experiments were carried out by triplicates, and the standard deviations were calculated by Excel. See also Figure S2.

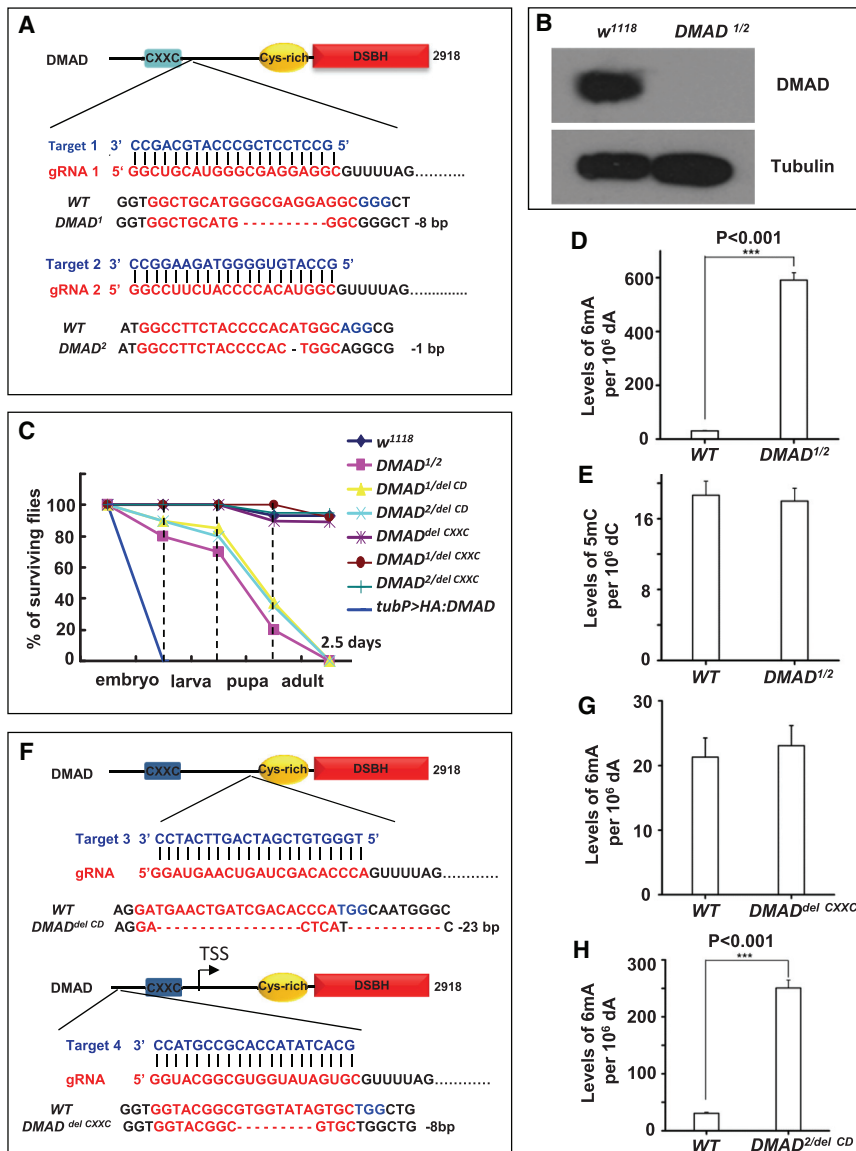


Figure 3. DMAD Is Required for *Drosophila* Development

(A) Schematic representation of *DMAD* mutant allele generation using the CRISPR/Cas system. The primer sequences of sgRNAs and information for *DMAD*¹ and *DMAD*² are indicated.

(B) Western blot experiments showed that the *DMAD* protein was completely abolished in the *DMAD*¹ and *DMAD*² allelic backgrounds.

(C) Survival rates of wild-type, different *DMAD* mutant, or overexpression flies at different developmental stages as indicated were measured.

(D and E) Abundance of 6mA (D) and 5mC (E) in DNA from wild-type and *DMAD*^{1/2} mutant flies was measured by mass spectrometry.

(F) Schematic representation for the generation of mutant alleles of *DMAD* using the CRISPR/Cas system. The primer sequences of sgRNAs and information for *DMAD*^{del-CXXC} and *DMAD*^{del-CD} are provided.

(G and H) Abundance of 6mA in DNA from wild-type, *DMAD*^{del-CXXC} (G) and *DMAD*^{2/del-CD} (H) mutant flies was measured by mass spectrometry. The experiments were carried out by triplicates, and the standard deviations were calculated by Excel. See also Figure S3.

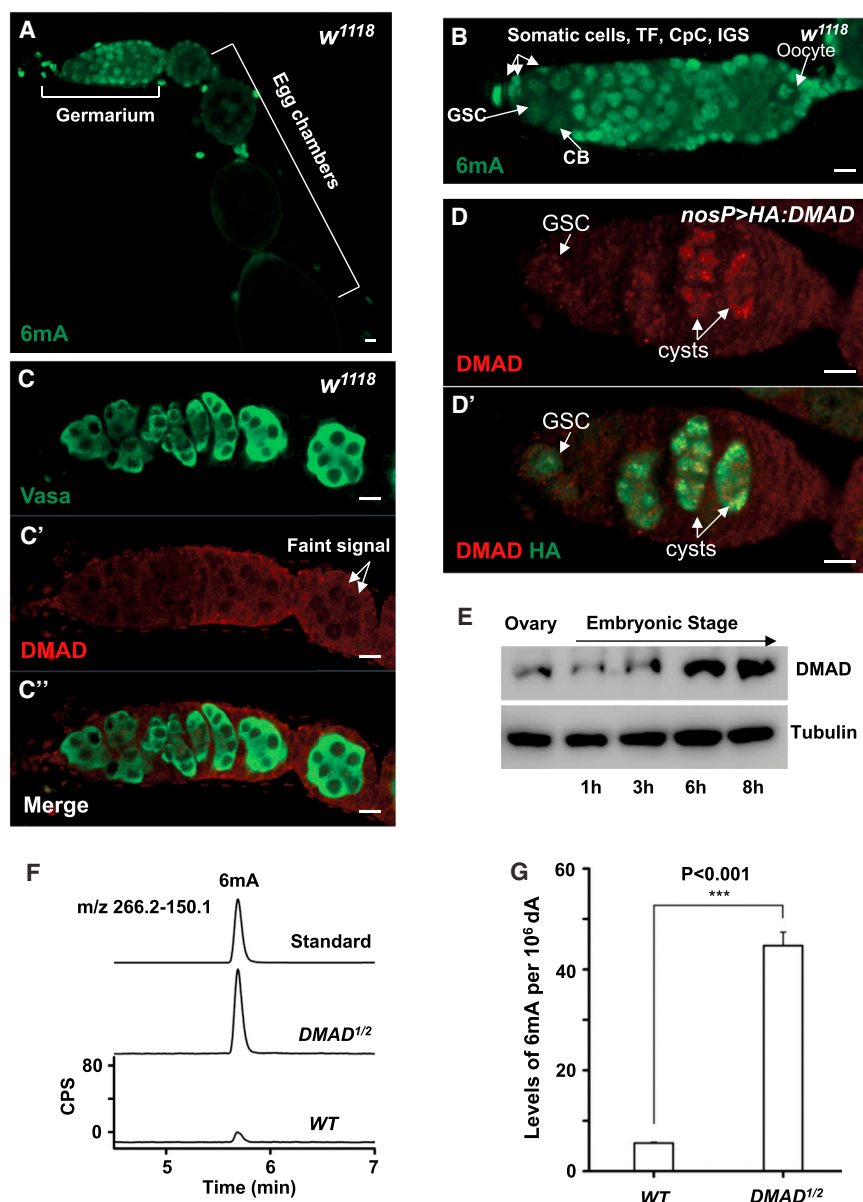
role of *DMAD* in development and in suppressing 6mA modification. By contrast, the *DMAD*^{del-CD} completely failed to complement either the *DMAD*¹ or *DMAD*² allele. The trans-heterozygous mutant *DMAD*^{del-CD}/*DMAD*² and *DMAD*^{del-CD}/*DMAD*¹ displayed strong developmental defects (Figure 3C). Interestingly, we found that the levels of 6mA modification were also increased in the trans-heterozygous mutant background that carried *DMAD*^{del-CD}, compared to wild-type (Figure 3H). These results together suggested that the catalytic domain is essential for the role of *DMAD* in development and in suppressing 6mA modification in vivo.

Additionally, we also examined the phenotypes in animals with ectopic expression of the *DMAD* by generating the transgenic flies, *P{UASp-HA:DMAD}*, in which the HA-tagged full-length *DMAD* was placed under the control of the UASp promoter. As shown in Figure 3C, overexpression of *DMAD* by the ubiquitous

driver, *tub-gal4*, at 29°C caused lethality at the late embryonic stage, since embryos (n = 735) expressing *DMAD* completely failed to develop to the larva stage. However, relative low levels of *DMAD* expression by the Gal4/Gal80ts system (data not shown) permitted ~28% (n = 810) of the *DMAD* expression embryos to develop to the larva stage. Interestingly, when we induced the expression of the *DMAD* at the 10 hr embryonic stage by taking advantage of the temperature-dependent activity of Gal80ts (see Extended Experimental Procedures), we found that ~49% of *DMAD*-expressing embryos (n = 530) could develop to the larva stage. Thus, our findings suggested that the *DMAD* expression must be under tight control during embryonic development.

DMAD Promotes Differentiation of Early Germ Cells in *Drosophila*

We next explored the potential roles of the *DMAD* in tissue homeostasis. The *Drosophila* ovary offers an excellent model system to study a number of important biological processes, such as germline stem cell (GSC) regulation, oocyte determination, and epigenetic control (Lin, 2002; Ohlstein et al., 2004; Spradling et al., 2001). A wild-type female contains a pair of ovaries, each of which is composed of 16–20 ovarioles that consist of an anterior functional unit (called “germarium”) and a linear string of differentiated egg chambers (Figures S4A and S4B). In the tip of germarium, GSCs divide asymmetrically to



produce two daughters. The anterior daughter cell retains contact with the cap cells as a new stem cell, whereas the posterior differentiating daughter cell becomes a cystoblast (CB) (Figures S4C and S4D). The CB further divides four times with incomplete cytokinesis, resulting in a cyst that sustains oogenesis (Figure S4D). To address whether DMAD has a role in germline, we performed immunostaining experiments to investigate the patterns of 6mA in the ovary. As shown in Figures 4A and 4B, a striking 6mA staining signal was detected in the nucleus of germarium cells, including germ cells and somatic cells (Figure 4B). In contrast, the 6mA signal was gradually reduced with development and ultimately disappeared in germ cells of mature differentiated egg chambers (Figure 4A), suggesting that 6mA modification occurs in the germ cell in a developmentally regulated fashion. We then determined DMAD expression

in the germarium using the anti-DMAD antibody. As shown in Figures 4C–4C'', we detected no nuclear staining of the DMAD in germarium germ cells, but a faint signal was present in the nucleus in egg chambers. To test whether the faint signal from the anti-DMAD antibody was specific, we performed further immunostaining in the germarium for ectopic expression of the HA-tagged DMAD. As shown in Figures 4D–4D' and S4E–S4E'', overlapping staining signals of HA with DMAD in the nucleus of germ cells in both germaria and egg chambers were readily detected. These findings together

suggest that DMAD expression occurs at a low level in the ovary. In support of this, our western blot assays showed that the DMAD protein expression was maintained at low levels in the ovary as compared with in the embryo (Figure 4E). We then tested whether DMAD has a role in regulating 6mA modification in ovaries and found that loss of *DMAD* resulted in an ~10-fold increase in levels of 6mA in ovaries (Figures 4F and 4G).

To test whether DMAD-mediated 6mA modification has a role in regulating early germ cell development, we examined the germ cell behavior in *DMAD* mutant ovaries by performing immunostaining assays using anti-Vasa and anti-Hts antibodies, which were used to visualize germ cells and fusomes, respectively. As shown in Figures 5A–5E, a newly enclosed (1-day-old) wild-type germarium normally contained 3–4 GSCs/CBs,

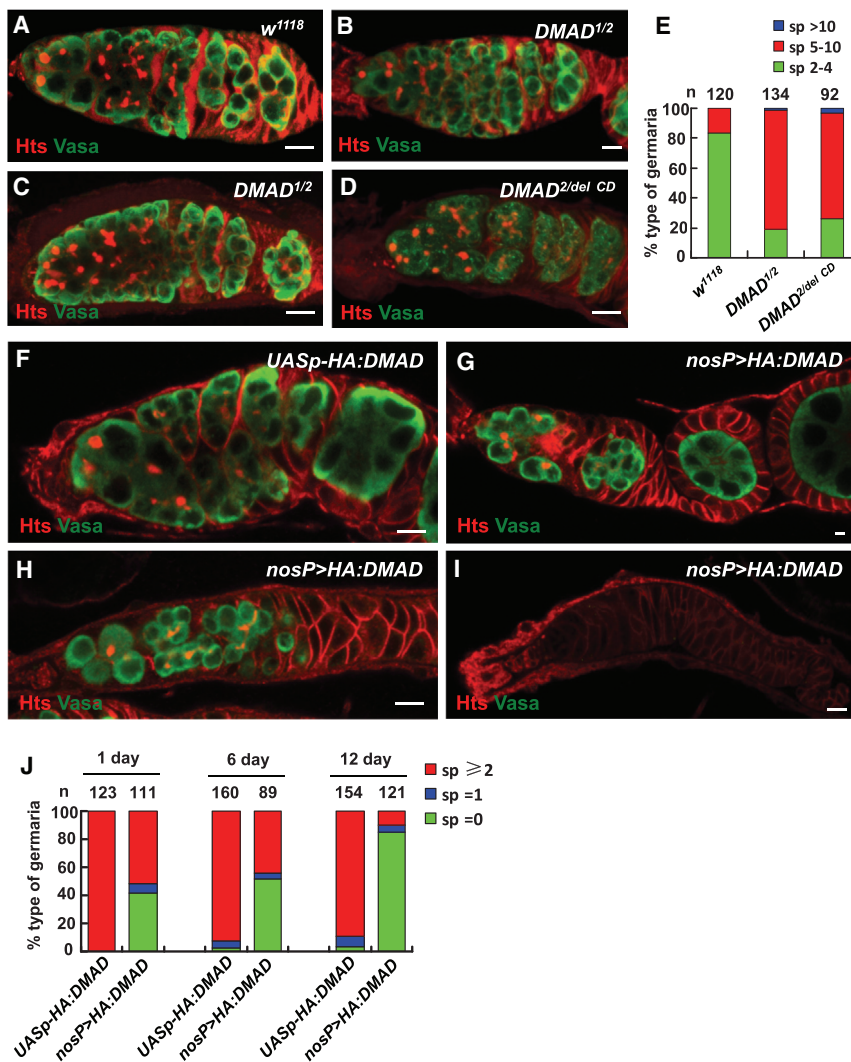


Figure 5. DMAD Promotes Early Germ Cell Differentiation

(A–D) Ovaries from wild-type (A) and different *DMAD* mutant flies as indicated were stained with anti-Hts (Red) and anti-Vasa (Green) antibodies. Scale bar, 10 μ m.

(E) Quantification assay showing percentage of type of germaria in wild-type and different *DMAD* mutant ovaries corresponding to (A–D). The types of germaria were classified according to the number of spectrosome (Sp)-containing germ cells in each germarium.

(F–I) Ovaries from wild-type and *P{nosP-gal4:vp16}/P{UASp-HA:DMAD}* transgenic flies were stained with anti-Hts (Red) and anti-Vasa (Green) antibodies. Scale bar, 10 μ m.

(J) Quantification assay showing percentage of type of germaria in wild-type and *P{nosP-gal4:vp16}/P{UASp-HA:DMAD}* ovaries corresponding to (F–I).

See also Figure S5.

support the *in vitro* 6mA demethylation reaction. In contrast, nuclear extracts from *DMAD* mutant ovaries with addition of the purified *DMAD*-CD protein, but not *DMAD*-CD^{mut} protein, resulted in striking enzymatic activity for 6mA demethylation (Figures 6A and 6B). As shown in Figure 6B, ~46% of 6mA bases in the substrates were demethylated, compared with only ~20% having demethylated 6mA in the control reaction with the addition of AlkB or with nuclear extracts from wild-type ovaries. Thus, our findings support that *DMAD* is essential for 6mA demethylation in *Drosophila*.

whereas the 1-day-old *DMAD* mutant had an average of >8 spectrosome-containing germ cells (GSC-like cells) (Figures 5C and 5D), suggesting that *DMAD* plays a role in promoting early germ cell differentiation. We next overexpressed *DMAD* to examine the phenotype in germ cells by generating flies carrying a transgene combination, *P{UASp-HA:DMAD}* and *P{nosP-gal4:vp16}*, in which *nosP-gal4:vp16* is a germ-cell-specific driver. As shown in Figures 5F–5J, S5A, and S5B, overexpression of *DMAD* led to a significant loss of germ cells, including GSCs, supporting that *DMAD* plays a role in promoting GSC differentiation.

DMAD Directly Catalyzes Demethylation of 6mA

To elucidate the biochemical properties of *DMAD*, we asked whether *DMAD* directly catalyzes 6mA demethylation by performing *in vitro* demethylation activity assays using the ovarian nuclear extracts from wild-type and *DMAD* mutants. As shown in Figure 6B, while wild-type ovarian nuclear extract has considerable enzymatic activity for 6mA demethylation, *DMAD* mutant nuclear extracts almost completely failed to

We next tested whether *DMAD* has a similar role in other tissues. The fly brain represents another interesting and complementary model to study the *DMAD*-mediated 6mA modification due to two reasons. First, *DMAD* is expressed at a much higher level in the fly brains than in ovaries (Figures 6C and 6D). Second, levels of 6mA are also relatively low in the brain when compared with very early-stage embryos (Figures 1C and 1E). We measured abundance of 6mA in brain genome from wild-type and *DMAD* mutant brain tissue, respectively. Strikingly, we found that loss of *DMAD* resulted in up to a 100-fold increase in 6mA levels in brain (Figures 6E and 6F). Additionally, similar to ovary, nuclear extracts from *DMAD* mutant brain with addition of the purified *DMAD*-CD protein exhibited a considerable 6mA demethylation activity (Figure S6A). Collectively, our findings reveal that *DMAD* plays a critical role in the regulation of 6mA demethylation in *Drosophila*.

Up to 100-fold increases of 6mA abundance in *DMAD* mutant tissues raised a possibility that potential 6mA methylases catalyze 6mA methylation in fly DNA. To explore this issue, we employed nuclear extracts from wild-type and

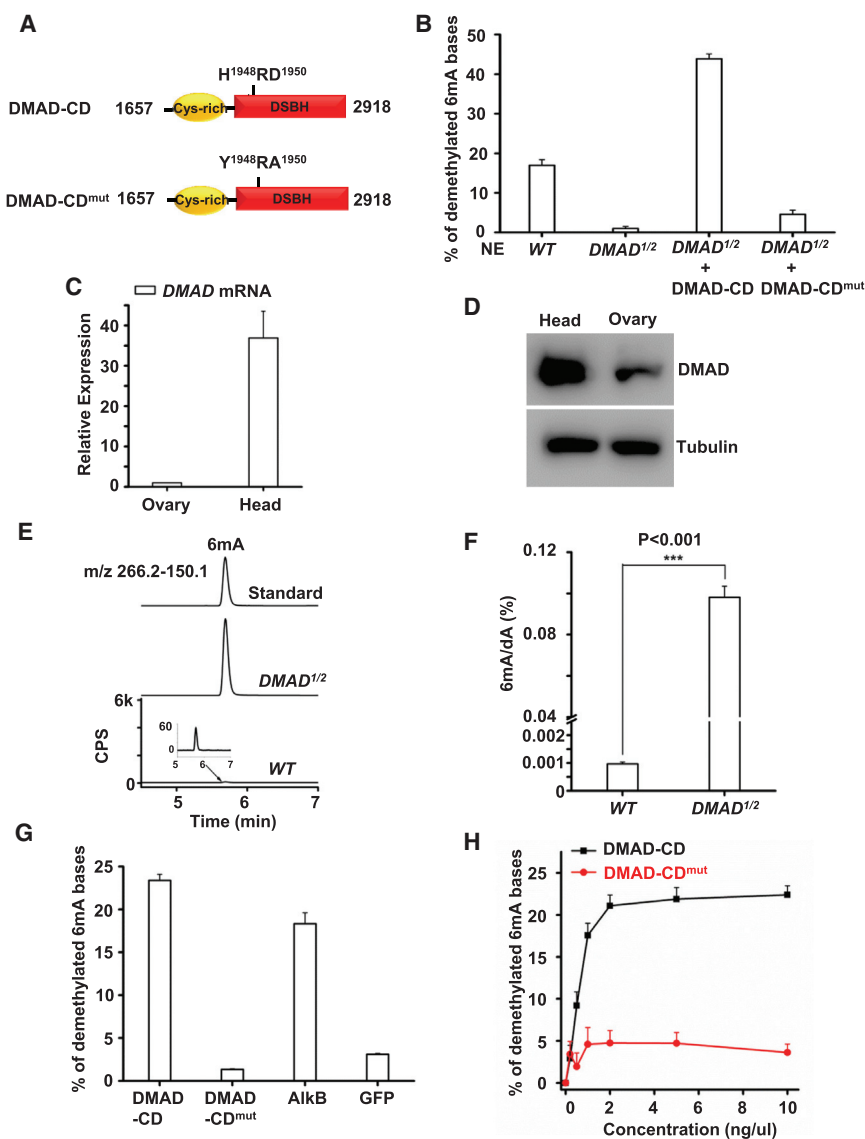


Figure 6. DMAD Directly Catalyzes Demethylation of 6mA

(A) Schematic diagram of the DMAD catalytic domain (DMAD-CD) fragment (aa. 1657–2918) and its mutant (DMAD-CD^{mut}), in which two residues, H1948 and D1950, were mutated to Y and A, respectively. These two Fe(II)-binding sites are located in the highly conserved “H-R/K/Q-D” motif, which has been shown to be important for the catalytic activity in the family of AlkB-like Fe(II)/ α -ketoglutarate-dependent dioxygenases.

(B) The *in vitro* 6mA demethylation assays were performed to test enzymatic activity in wild-type ovary nuclear extracts, DMAD mutant ovary nuclear extracts without or with addition of the DMAD-CD, or DMAD-CD^{mut} protein as indicated. (C and D) Levels of DMAD mRNA (C) and protein expression (D) in ovary and brain were measured by qRT-PCR and western blot assays, respectively.

(E and F) UHPLC-MRM-MS/MS chromatograms (E) and quantification (F) showing 6mA abundance in genomic DNA from wild-type and DMAD^{1/2} mutant brains.

(G) Comparison for the enzymatic activity of DMAD with its CD mutant form, AlkB, in the *in vitro* 6mA demethylation assays.

(H) An *in vitro* enzymatic assay showing that the DMAD protein directly catalyzes the 6mA demethylation in a concentration-dependent manner.

The experiments were carried out by triplicates, and the standard deviations were calculated by Excel. See also Figure S6.

DMAD-Mediated 6mA Demethylation Is Correlated with Transposon Expression

We next sought to test whether DMAD influences 6mA modification of the *Drosophila* genome. We collected genomic DNA from 1-day-old wild-type and DMAD mutant ovaries and performed DNA immunoprecipitation (DNA-IP) experiments using anti-6mA antibody

DMAD mutant flies to perform *in vitro* 6mA methylation assays. As shown in Figures S6B and S6C, wild-type nuclear extracts showed a weak enzymatic activity for 6mA methylation, and DMAD mutant nuclear extracts exhibited relatively high 6mA methylation activities. These findings suggest that potential 6mA methylases and DMAD constitute an antagonistic loop to control 6mA base modification. Nevertheless, our findings suggest that demethylation activity of DMAD plays a predominant role in maintaining low levels of 6mA in genome. To determine whether DMAD directly participates in 6mA demethylation, we performed *in vitro* DNA demethylation assays. As shown in Figures 6G, 6H, and S6D, the purified catalytic domain of DMAD (DMAD-CD), but not its dead form of DMAD (DMAD-CD^{mut}), is sufficient to promote 6mA demethylation, suggesting that DMAD is the *Drosophila* 6mA demethylase.

and then generated DNA libraries, which were subjected to a high-throughput sequencing analysis. In this assay, the IgG-immunoprecipitated DNA from an equivalent amount of wild-type ovaries was used as the control, and 4.2–5.5 million reads were obtained through high-throughput sequencing. We then used MACS software (2.0 version, Zhang et al., 2008) to identify the 6mA-enriched regions. In sum, we identified 161 and 491 peaks from wild-type and DMAD mutant samples, respectively (Figure 7A). 88% of peaks identified in wild-type are also identified in DMAD mutant sample, while 73% of peaks in DMAD mutant sample were unique (Figure 7A). As shown in Figures 7B and 7C, signal of 6mA was stronger in DMAD mutant sample than wild-type sample with respect to both common peaks and DMAD mutant unique peaks, providing further evidence that 6mA demethylation is regulated by DMAD.

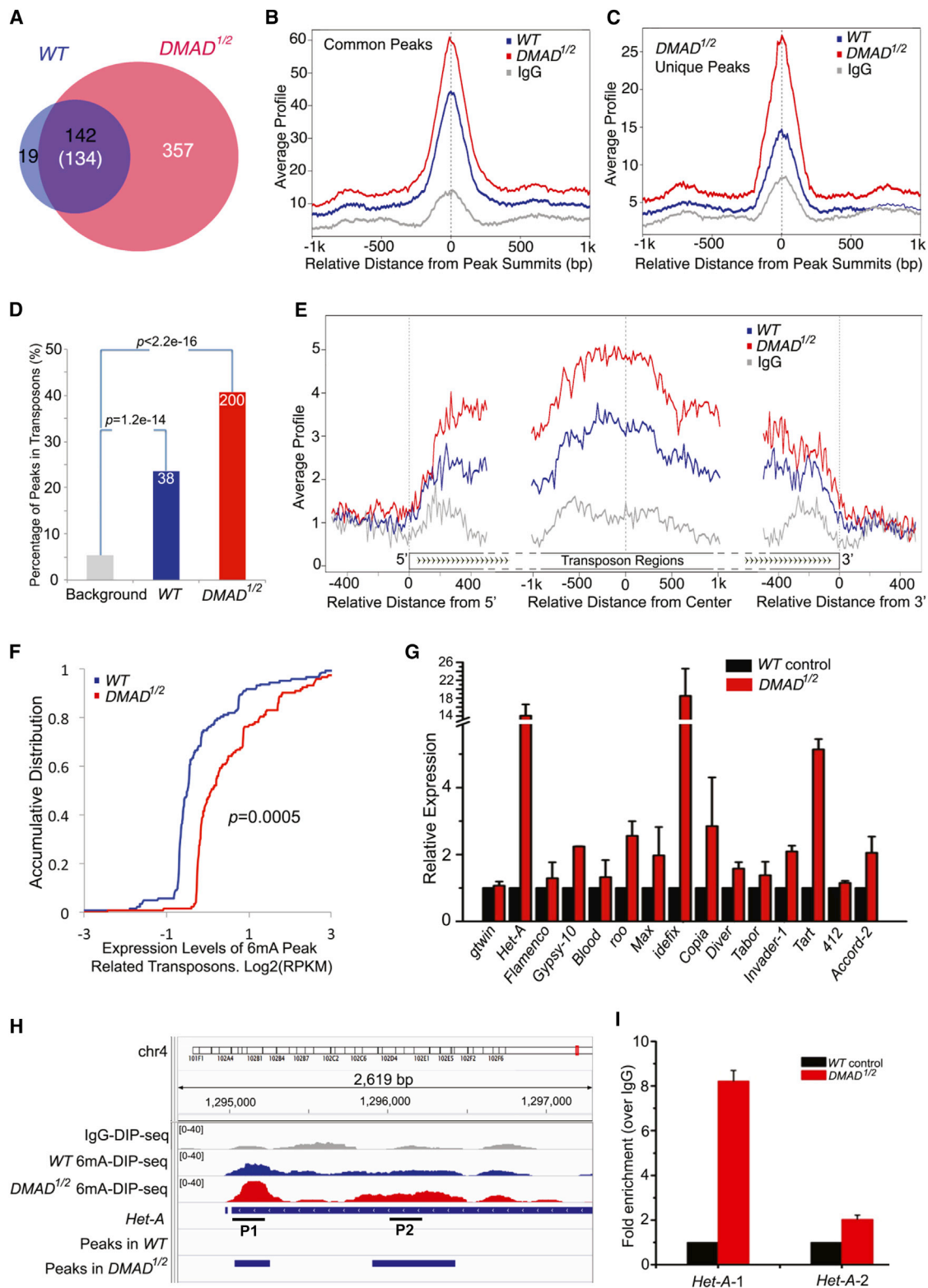


Figure 7. DMAD Controls 6mA Modification on Genome and Transposon Silencing

(A) 6mA enrichment peaks identified from wild-type and *DMAD* mutant ovary samples. A significant portion of peaks are shared by wild-type and *DMAD* mutant ovary samples.

(legend continued on next page)

Of note, our immunostaining assays revealed an evident expansion of γ -H2Av expression domain in late meiosis in *DMAD* mutant germaria, compared with the wild-type control (Figures S5C–S5D'). This phenomenon is also present in *ago3* mutant germaria (Huang et al., 2014). We thus reasoned that *DMAD* might be involved in transposon silencing by regulating 6mA modification. Indeed, we found that 24% and 41% of peaks from wild-type and *DMAD* mutants are located in the transposon regions, respectively (Figure 7D), indicating that the transposon sequence is the important 6mA-modified target by *DMAD* in the genome. Additionally, we found that the 6mA signal was much more enriched in the gene body of transposons when compared with the upstream and downstream regions (Figure 7E). To link the 6mA modification with transposon expression, we employed wild-type and *DMAD* mutant samples to perform RNA-seq analysis. Global expression profiling analysis revealed that transposons with 6mA peaks express significantly higher in *DMAD* mutant ovary than do those in wild-type (Figure 7F).

We next performed qRT-PCR assays and measured levels of transposon transcripts in wild-type and *DMAD* mutant ovaries, respectively. As shown in Figure 7G, loss of *DMAD* led to an increase in levels of most of the transposon transcripts that we chose to evaluate. Particularly, *Idefix*, *Het-A*, *Tart*, and *Copia* were significantly increased in the *DMAD* mutant, compared with the control. Importantly, the 6mA DNA immunoprecipitation followed by qPCR assays further confirmed that more 6mA modification occurs on the transposon regions in *DMAD* mutant ovaries than in wild-type ovaries (Figures 7H, 7I, and S7A–S7D). Taken together, our findings suggest that *DMAD*-mediated 6mA demethylation is correlated with transposon expression.

DISCUSSION

In this study, we find that 6mA is present in the *Drosophila* DNA at a relatively high level at the very earliest embryonic stages but at low levels at the late embryonic stages. Moreover, we show that the dynamic change of 6mA modification during embryonic development may involve an active demethylation event, a process that is primarily regulated by the *Drosophila* *DMAD* protein. Importantly, *DMAD* is essential for *Drosophila* development and tissue homeostasis, perhaps partially by suppressing adenine methylation and transposon expression in ovary. Thus, our study suggests a potential role

of the *DMAD*-6mA regulatory axis in controlling development in higher eukaryotes.

Adenine Methylation in *Drosophila*

To date, studies examining 6mA as a biologically relevant, methylated DNA base have mainly been limited to bacteria, although it is well known that 6mA is also present in the genomic DNA of several unicellular eukaryotes (Wion and Casadesús, 2006). Previous studies have reported that, while 5mC is enriched in genomes of many higher eukaryotes, particularly mammals, a signal for 6mA has not been detected (Ratel et al., 2006). Because the important function of 5mC modification in mammals has attracted much interest in the field of the epigenetic control, the issue of whether adenine methylation occurs in general and its related roles in higher eukaryote DNA has remained largely unresolved. The previous failure to detect 6mA in higher-eukaryote DNA could be that its level is so low in eukaryotes that it was undetectable with the technology used in previous reports (Lawley et al., 1972; Vanyushin et al., 1970). However, adenine methylation might occur in a tissue-specific or in a developmentally regulated manner in higher-eukaryote cells (Raddatz et al., 2013), and low levels of 6mA could be controlled by a tight negative regulatory mechanism mediated by 6mA demethylases. Thus, searching for specific 6mA demethylases is important for understanding the potential role of adenine methylation in higher-eukaryote cells. In this study, we show that the *Drosophila* *DMAD* directly catalyzes 6mA demethylation in our biochemical assays, suggesting that it is likely an 6mA demethylase. Moreover, our functional assays show that loss of *DMAD* leads to strong developmental defects and significantly increases the abundance of 6mA modification in DNA. These findings bring an insight into understanding the potential function of 6mA modification in development and tissue homeostasis in higher eukaryotes. In this study, we find that *DMAD*-mediated 6mA demethylation is correlated with transposon suppression in ovary, indicating that 6mA modification as an epigenetic mark likely regulates gene expression. However, the possibility of other *DMAD*-mediated processes contributing to normal development cannot be completely ruled out and warrants further investigation.

The discovery that loss of *DMAD* causes a dramatic increase of 6mA abundance in adult tissues opens an interesting possibility of the existence of potential 6mA methylases in flies for DNA 6mA methylation. Our in vitro enzymatic assays revealed that *Drosophila* nuclear extracts possess both methylation and

(B and C) The average 6mA signal strength of all common peaks (B) and *DMAD* mutant unique peaks (C). 6mA signal was stronger in *DMAD* ovary mutant sample than in wild-type ovary sample.

(D) Percentage of 6mA enrichment peaks located in transposon regions. 6mA peaks were significantly located in transposon regions.

(E) The average 6mA signal strength on all 6mA peak-related transposons. The 6mA signal was enriched in transposon body.

(F) Accumulative distribution of expression levels of 6mA peak-related transposons in wild-type and *DMAD* mutant ovary samples. The p value represent Wilcoxon rank sum test.

(G) qRT-PCR experiments were performed to measure the transposon expression levels in wild-type control and *DMAD* mutant ovary.

(H) The 6mA-enriched regions are found in the *Het-A* transposon region in the indicated chromosome, in which *DMAD* mutant samples show higher enrichment when compared with that of wild-type samples.

(I) qPCR experiments were performed to confirm the 6mA-enriched regions indicated in (H) when *DMAD* mutant samples were compared with that of wild-type. In this assay, the corresponding regions IPed by IgG were used for normalization. The experiments were carried out by triplicates, and the standard deviations were calculated by Excel. See also Figure S7.

demethylation enzymatic activities for 6mA base modification. Thus, it is likely that the potential 6mA methylase(s) and DMAD act antagonistically to maintain the proper modification of 6mA in flies. It would be interesting to identify specific 6mA methylases in the future.

Another question is whether DMAD is involved in the DNA damage process and repair DNA methylation lesions. In this study, we have measured the levels of 1mA, 3mC, 3mA, and m6G because 1mA and 3mC are predominant forms of base damage in single-stranded DNA, and 3mA and m6G are products in double-stranded DNA damage (Lindahl et al., 1988; Treweek et al., 2002). Our results revealed that loss of DMAD did not cause apparent increase in levels of these bases. Additionally, methyl iodide treatment did not cause apparent upregulation of 6mA levels but led to a dramatic increase in levels of m6G and 3mC in fly genomic DNA (Figures S3G–S3I). Thus, our findings strongly argue that 6mA comes from enzymatic installation rather than as a byproduct of DNA damage.

The Role of DMAD in 6mA Demethylation in *Drosophila*

The controversy over 6mA in mammalian DNA is similar to that of 5mC modification in *Drosophila* and has been discussed for a long time (Raddatz et al., 2013). A recent work suggested that the *Drosophila* genome lacks a defined 5mC pattern (Raddatz et al., 2013). In this study, we find that no evidence supports oxidation of 5mC in *Drosophila*. Although DMAD can catalyze the 5mC oxidation in in vitro enzymatic reactions, the in vivo studies revealed no difference in levels of 5mC and 5hmC detected in DNA from the wild-type and DMAD mutant flies, revealing that the DMAD has no role in catalyzing 5mC oxidation in vivo.

A broadly accepted concept is that DNA base modification through methylation plays evolutionarily conserved epigenetic roles in a wide array of organisms from bacteria to animals, although the underlying mechanisms might be different among species (Wion and Casadesús, 2006). From an evolutionary perspective, since its DNA is not methylated at cytosine, *Drosophila* likely uses other types of methylated bases, such as 6mA, to fulfill the function of 5mC in mammals. The discovery that DMAD possesses enzymatic activity for 6mA demethylation, as well as the identification of its role in suppressing 6mA modification in vivo, support that DMAD functions as an 6mA demethylase in *Drosophila*.

All AlkB family members contain a core catalytic domain called the double-stranded β -helix (DSBH) fold (Shen et al., 2014). Our results suggest that the DSBH domain in DMAD is essential for its function in regulating 6mA demethylation in flies. It would be interesting to search for DSBH-domain-containing dioxygenases responsible for 6mA demethylation in mammals in the future. Members of Tet family proteins, without a doubt, are attractive candidates.

EXPERIMENTAL PROCEDURES

Drosophila Strains

Fly stocks used in this study were maintained under standard culture conditions. The *w¹¹¹⁸* strain was used as the host for all P-element-mediated transpositions. Strains *P{tubP-gal80ts}*, *P{tubP-gal4}*, and *P{nosP-gal4:vp16}*

have been maintained in the Chen lab. *P{UASp-HA:DMAD}* was generated in this study; *DMAD¹*, *DMAD²*, *DMAD^{del-CXXC}*, and *DMAD^{del-CD}* were generated by the CRISPR/Cas system in this study. See the [Extended Experimental Procedures](#) for a more detailed protocol for generation of DMAD null alleles using CRISPR/Cas system.

Immunohistochemistry

Ovaries were prepared for immunohistochemistry as described previously (Yang et al., 2007). See the [Extended Experimental Procedures](#) for a more detailed protocol.

Gene Knockdown in *Drosophila* Embryos

The dsRNA fragments corresponding to DMAD and *gfp* mRNAs were synthesized in a PCR reaction and then fused to the T7 RNA polymerase binding site at both 5' and 3' ends, which were used to generate the DMAD and *gfp* dsRNA in vitro by using the RiboMAX Large Scale RNA kit (Promega) following the manufacturer's instructions. The DMAD or *gfp* dsRNA (1 μ g/ μ l) was injected into *w¹¹¹⁸* embryos. The embryos were incubated at room temperature for turnover of the target protein.

Purification of Nuclear Extracts and Genomic DNA

Nuclear extracts were extracted from embryos or adult tissues using Minute Cytoplasmic and Nuclear Extraction Kit (Invent Biotech). Genomic DNA was extracted with Wizard genomic DNA purification Kit (Promega) following the manufacturer's instructions.

Anti-DMAD Antibodies

The anti-DMAD antibody was generated by immunizing rabbit and mouse with the recombinant protein GST-DMAD (amino acids 959–1108) produced in *E. coli*.

Immunodepletion Experiments

For immunodepletion experiments, 10 μ l of protein A/G beads were mixed with 200 μ l of hypotonic buffer (plus 0.02% CHAPS, 0.1 mM PMSF). This solution was mixed with 10 μ g rabbit anti-DMAD antibody or rabbit IgG and was rotated for about 1 hr using a head-to-tail roller at 4°C. Embryonic nuclear extracts were obtained as described above and were subjected to immunoprecipitation using protein A/G beads treated with antibodies at 4°C for 2 hr. Subsequently, samples were centrifuged and supernatants were collected and used for in vitro 6mA demethylation assays.

Quantitative Real-Time PCR Analysis

qRT-PCR experiments were performed as described previously (Huang et al., 2014). See the [Extended Experimental Procedures](#) for a more detailed protocol.

Dot Blot Assay

Different amounts of standard DNA either containing the base dA or 6mA and fly genomic DNA were used for dot blot assay. See the [Extended Experimental Procedures](#) for a more detailed protocol.

UHPLC-MRM-MS/MS Analysis

Genomic DNA was enzymatically digested into single nucleosides by a mixture of DNaseI, calf intestinal phosphatase, and snake venom phosphodiesterase I at 37°C for 12 hr. After the enzymes were removed by ultrafiltration, the digested DNA was subjected to UHPLC-MS/MS analysis (Yin et al., 2013). HPLC fractionation of *Drosophila* 6mA and UHPLC-QTOF-MS/MS analysis are shown in the [Extended Experimental Procedures](#) with a more detailed protocol.

In Vitro 6mA Demethylation

Calf thymus (CT) dsDNA was methylated by Dam methyltransferase following the manufacturer's instructions. The detailed protocol for 6mA demethylation can be seen in the [Extended Experimental Procedures](#).

Statistical Analysis

Data are presented as the mean \pm SEM from at least three independent experiments. Student's *t* test was used for comparison of two independent groups. For all tests, a *p* < 0.05 was considered statistically significant.

ACCESSION NUMBERS

All sequencing data are available at NCBI SRA under accession number SRA: SRP055483.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures and seven figures and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.04.018>.

AUTHOR CONTRIBUTIONS

G.Z. generated reagents (KO and transgenic flies and antibodies) and performed genetic and molecular biology experiments. H.H. and G.Z. performed enzymatic assays, and H.H. performed all mass-spec analysis. D.L. and G.Z. constructed libraries for the analysis of DNA-IP-seq and RNA-seq. S.H. performed all bioinformatics.

ACKNOWLEDGMENTS

This work is supported by the National Basic Research Program of China (2013CB945002), the Strategic Priority Research Program of the Chinese Academy of Sciences (XDB14030200 and XDA01010306), Natural Science Foundation of China (31130036, 21125523, 21327006, and 91019022), and the Chinese Academy of Sciences (YSW2013A01).

Received: December 22, 2014

Revised: February 25, 2015

Accepted: March 25, 2015

Published: April 30, 2015

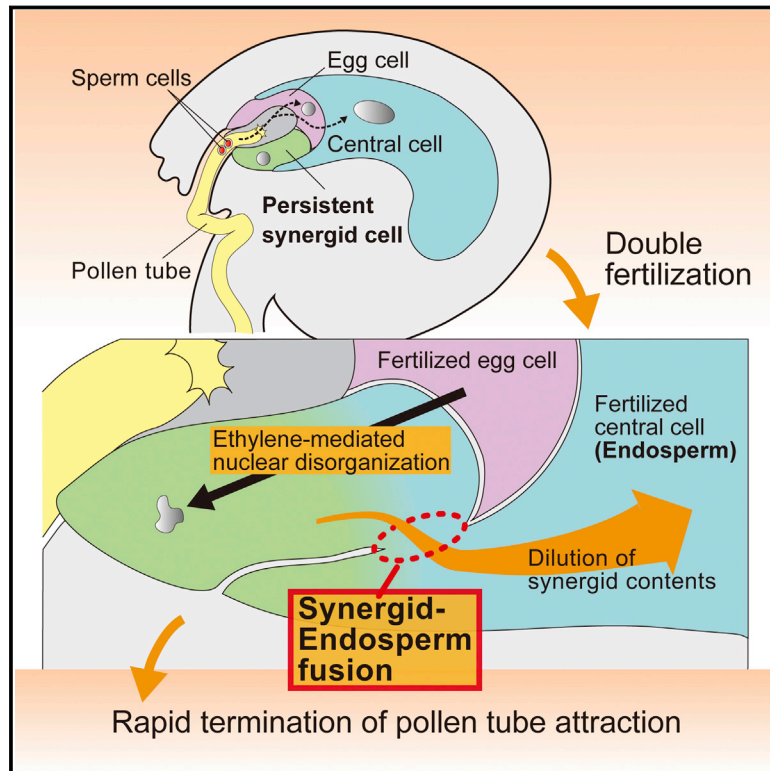
REFERENCES

- Bhutani, N., Burns, D.M., and Blau, H.M. (2011). DNA demethylation dynamics. *Cell* 146, 866–872.
- Cheng, X. (1995). Structure and function of DNA methyltransferases. *Annu. Rev. Biophys. Biomol. Struct.* 24, 293–318.
- Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., and Zhang, F. (2013). Multiplex genome engineering using CRISPR/Cas systems. *Science* 339, 819–823.
- Feng, S., Jacobsen, S.E., and Reik, W. (2010). Epigenetic reprogramming in plant and animal development. *Science* 330, 622–627.
- He, Y.F., Li, B.Z., Li, Z., Liu, P., Wang, Y., Tang, Q., Ding, J., Jia, Y., Chen, Z., Li, L., et al. (2011). Tet-mediated formation of 5-carboxylcytosine and its excision by TDG in mammalian DNA. *Science* 333, 1303–1307.
- Huang, H., Li, Y., Szulwach, K.E., Zhang, G., Jin, P., and Chen, D. (2014). AGO3 Slicer activity regulates mitochondria-nuage localization of Armitage and piRNA amplification. *J. Cell Biol.* 206, 217–230.
- Ito, S., Shen, L., Dai, Q., Wu, S.C., Collins, L.B., Swenberg, J.A., He, C., and Zhang, Y. (2011). Tet proteins can convert 5-methylcytosine to 5-formylcytosine and 5-carboxylcytosine. *Science* 333, 1300–1303.
- Julio, S.M., Heithoff, D.M., Provenzano, D., Klose, K.E., Sinsheimer, R.L., Low, D.A., and Mahan, M.J. (2001). DNA adenine methylase is essential for viability and plays a role in the pathogenesis of *Yersinia pseudotuberculosis* and *Vibrio cholerae*. *Infect. Immun.* 69, 7610–7615.
- Kriaucionis, S., and Heintz, N. (2009). The nuclear DNA base 5-hydroxymethylcytosine is present in Purkinje neurons and the brain. *Science* 324, 929–930.
- Lawley, P.D., Crathorn, A.R., Shah, S.A., and Smith, B.A. (1972). Bio-methylation of deoxyribonucleic acid in cultured human tumour cells (HeLa). Methylated bases other than 5-methylcytosine not detected. *Biochem. J.* 128, 133–138.
- Li, D., Delaney, J.C., Page, C.M., Yang, X., Chen, A.S., Wong, C., Drennan, C.L., and Essigmann, J.M. (2012). Exocyclic carbons adjacent to the N6 of adenine are targets for oxidation by the *Escherichia coli* adaptive response protein AlkB. *J. Am. Chem. Soc.* 134, 8896–8901.
- Lin, H. (2002). The stem-cell niche theory: lessons from flies. *Nat. Rev. Genet.* 3, 931–940.
- Lindahl, T., Sedgwick, B., Sekiguchi, M., and Nakabeppu, Y. (1988). Regulation and expression of the adaptive response to alkylating agents. *Annu. Rev. Biochem.* 57, 133–157.
- Lyko, F., and Maleszka, R. (2011). Insects as innovative models for functional studies of DNA methylation. *Trends Genet.* 27, 127–131.
- Lyko, F., Ramsahoye, B.H., and Jaenisch, R. (2000). DNA methylation in *Drosophila melanogaster*. *Nature* 408, 538–540.
- Maiti, A., and Drohat, A.C. (2011). Thymine DNA glycosylase can rapidly excise 5-formylcytosine and 5-carboxylcytosine: potential implications for active demethylation of CpG sites. *J. Biol. Chem.* 286, 35334–35338.
- Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. *Science* 339, 823–826.
- Ohlstein, B., Kai, T., Decotto, E., and Spradling, A. (2004). The stem cell niche: theme and variations. *Curr. Opin. Cell Biol.* 16, 693–699.
- Raddatz, G., Guzzardo, P.M., Olova, N., Fantappiè, M.R., Rampp, M., Schaefer, M., Reik, W., Hannon, G.J., and Lyko, F. (2013). Dnmt2-dependent methylomes lack defined DNA methylation patterns. *Proc. Natl. Acad. Sci. USA* 110, 8627–8631.
- Ratel, D., Ravanat, J.L., Berger, F., and Wion, D. (2006). N6-methyladenine: the other methylated base of DNA. *BioEssays* 28, 309–315.
- Reisenauer, A., Kahng, L.S., McCollum, S., and Shapiro, L. (1999). Bacterial DNA methylation: a cell cycle regulator? *J. Bacteriol.* 181, 5135–5139.
- Shen, L., Song, C.X., He, C., and Zhang, Y. (2014). Mechanism and function of oxidative reversal of DNA and RNA methylation. *Annu. Rev. Biochem.* 83, 585–614.
- Smith, Z.D., and Meissner, A. (2013). DNA methylation: roles in mammalian development. *Nat. Rev. Genet.* 14, 204–220.
- Spradling, A., Drummond-Barbosa, D., and Kai, T. (2001). Stem cells find their niche. *Nature* 414, 98–104.
- Stephens, C., Reisenauer, A., Wright, R., and Shapiro, L. (1996). A cell cycle-regulated bacterial DNA methyltransferase is essential for viability. *Proc. Natl. Acad. Sci. USA* 93, 1210–1214.
- Tahiliani, M., Koh, K.P., Shen, Y., Pastor, W.A., Bandukwala, H., Brudno, Y., Agarwal, S., Iyer, L.M., Liu, D.R., Aravind, L., and Rao, A. (2009). Conversion of 5-methylcytosine to 5-hydroxymethylcytosine in mammalian DNA by MLL partner TET1. *Science* 324, 930–935.
- Trewick, S.C., Henshaw, T.F., Hausinger, R.P., Lindahl, T., and Sedgwick, B. (2002). Oxidative demethylation by *Escherichia coli* AlkB directly reverts DNA base damage. *Nature* 419, 174–178.
- Vanyushin, B.F., Tkacheva, S.G., and Belozersky, A.N. (1970). Rare bases in animal DNA. *Nature* 225, 948–949.
- Wion, D., and Casadesús, J. (2006). N6-methyl-adenine: an epigenetic signal for DNA-protein interactions. *Nat. Rev. Microbiol.* 4, 183–192.
- Wright, R., Stephens, C., and Shapiro, L. (1997). The CcrM DNA methyltransferase is widespread in the alpha subdivision of proteobacteria, and its

- essential functions are conserved in *Rhizobium meliloti* and *Caulobacter crescentus*. *J. Bacteriol.* **179**, 5869–5877.
- Wu, H., and Zhang, Y. (2014). Reversing DNA methylation: mechanisms, genomics, and biological functions. *Cell* **156**, 45–68.
- Yang, L., Chen, D., Duan, R., Xia, L., Wang, J., Qurashi, A., Jin, P., and Chen, D. (2007). Argonaute 1 regulates the fate of germline stem cells in *Drosophila*. *Development* **134**, 4265–4272.
- Yin, R., Mao, S.Q., Zhao, B., Chong, Z., Yang, Y., Zhao, C., Zhang, D., Huang, H., Gao, J., Li, Z., et al. (2013). Ascorbic acid enhances Tet-mediated 5-methylcytosine oxidation and promotes DNA demethylation in mammals. *J. Am. Chem. Soc.* **135**, 10396–10403.
- Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nusbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* **9**, R137.

Rapid Elimination of the Persistent Synergid through a Cell Fusion Mechanism

Graphical Abstract



Authors

Daisuke Maruyama, Ronny Völz, ..., Rita Groß-Hardt, Tetsuya Higashiyama

Correspondence

maruyama.daisuke@f.mbox.nagoya-u.ac.jp

In Brief

Two female gametes (the egg cell and the central cell) in flowering plants coordinately prevent attractions of excess number of pollen tubes via two mechanisms to inactivate persistent synergid cell.

Highlights

- The persistent synergid cell fuses with the endosperm after fertilization
- Synergid cytoplasm containing pollen tube attractants is diluted by the fusion
- FIS-PRC2 is involved in mitosis-associated synergid nuclear elimination
- Each female gamete independently controls the cell-fusion and ethylene signaling



Rapid Elimination of the Persistent Synergid through a Cell Fusion Mechanism

Daisuke Maruyama,^{1,2,3,4,*} Ronny Völz,^{5,6} Hidenori Takeuchi,⁷ Toshiyuki Mori,⁸ Tomoko Igawa,⁹ Daisuke Kurihara,⁷ Tomokazu Kawashima,^{4,10} Minako Ueda,¹ Masaki Ito,¹¹ Masaaki Umeda,^{12,13} Shuh-ichi Nishikawa,¹⁴ Rita Groß-Hardt,^{5,15} and Tetsuya Higashiyama^{1,2,7}

¹Institute of Transformative Bio-Molecules (ITbM), Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8602, Japan

²Division of Biological Science, Graduate School of Science, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8602, Japan

³Institute for Advanced Research, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8602, Japan

⁴Temasek Life Sciences Laboratory and Department of Biological Sciences, National University of Singapore, Singapore 117604, Singapore

⁵Center for Plant Molecular Biology (ZMBP), University of Tübingen, 72076 Tübingen, Germany

⁶Center for Desert Agriculture, Division of Biological and Environmental Sciences and Engineering, King Abdullah University of Science and Technology, Thuwal 23955-6900, Saudi Arabia

⁷JST ERATO Higashiyama Live-Holomics Project, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8602, Japan

⁸Waseda Institute for Advanced Study, Waseda University, 1-6-1 Nishiwaseda, Shinjuku-ku, Tokyo 169-8050, Japan

⁹Graduate School of Horticulture, Chiba University, 648 Matsudo, Matsudo-City, Chiba 271-8510, Japan

¹⁰Gregor Mendel Institute, Dr-Bohr-Gasse 3, 1030 Vienna, Austria

¹¹Graduate School of Bioagricultural Sciences and School of Agricultural Sciences, Nagoya University, Furo-cho, Chikusa-ku, Nagoya, Aichi 464-8602, Japan

¹²Graduate School of Biological Sciences, Nara Institute of Science and Technology, 8916-5 Takayama, Ikoma, Nara 630-0192, Japan

¹³JST, CREST, 8916-5 Takayama, Ikoma, Nara 630-0192, Japan

¹⁴Department of Life and Food Science, Graduate School of Science, Niigata University, 8050, Ikarashi 2-no-cho, Nishi-ku, Niigata 950-2181, Japan

¹⁵Center for Biomolecular Interactions Bremen, University of Bremen, Leobener Straße NW2 28359, Germany

*Correspondence: maruyama.daisuke@f.mbox.nagoya-u.ac.jp

<http://dx.doi.org/10.1016/j.cell.2015.03.018>

SUMMARY

In flowering plants, fertilization-dependent degeneration of the persistent synergid cell ensures one-on-one pairings of male and female gametes. Here, we report that the fusion of the persistent synergid cell and the endosperm selectively inactivates the persistent synergid cell in *Arabidopsis thaliana*. The synergid-endosperm fusion causes rapid dilution of pre-secreted pollen tube attractant in the persistent synergid cell and selective disorganization of the synergid nucleus during the endosperm proliferation, preventing attractions of excess number of pollen tubes (polytubey). The synergid-endosperm fusion is induced by fertilization of the central cell, while the egg cell fertilization predominantly activates ethylene signaling, an inducer of the synergid nuclear disorganization. Therefore, two female gametes (the egg and the central cell) control independent pathways yet coordinately accomplish the elimination of the persistent synergid cell by double fertilization.

INTRODUCTION

During sexual reproduction, a female gamete must be fertilized by a single male gamete to generate diploid zygote. In animals, the egg has polyspermy block mechanisms that prevent multiple fertilizations by more than one sperm (Gardner and Evans, 2006;

Tsaadon et al., 2006). Flowering plants have similar system preventing additional gametic fusion (Scott et al., 2008). However, such situation rarely happens in vivo, because an ovule receives exactly two sperm cells for double fertilization; a single pollen tube delivers two sperm cells that independently fertilize the egg cell and the central cell to produce embryo and endosperm, respectively (Figure 1A) (Maheshwari, 1950). Attractions of excess number of pollen tubes (polytubey) are prevented by a mechanism recently defined as polytubey block (Beale et al., 2012; Beale and Johnson, 2013).

One of the central mechanisms of polytubey block is a cessation of pollen tube attraction and this attraction is precisely controlled by synergid cells (Takeuchi and Higashiyama, 2011) (Figure 1A). In most flowering plants including *Arabidopsis thaliana*, mature ovule contains seven-celled embryo sac consisting of two synergid cells, an egg cell, a central cell, and three antipodal cells (Figure 1A). The synergid cells have a characteristic invagination of cell wall facing toward the entrance of the pollen tube (micropyle). This invaginated structure termed filiform apparatus actively secretes peptides such as AtLURE1, a cysteine-rich peptide that is required and sufficient for pollen tube attraction in *A. thaliana*. (Takeuchi and Higashiyama, 2012). Upon successful fertilization, synergid cells are determined to die by either of two ways. When receiving pollen tube discharge, one synergid cell degenerates and is termed the degenerated synergid cell (Figure 1B). The other synergid cell, termed the persistent synergid cell, undergoes nuclear degeneration within a few hours after successful double fertilization (Figures 1C and 1D) (Beale et al., 2012; Völz et al., 2013). The consecutive synergid degenerations result in the cessation of

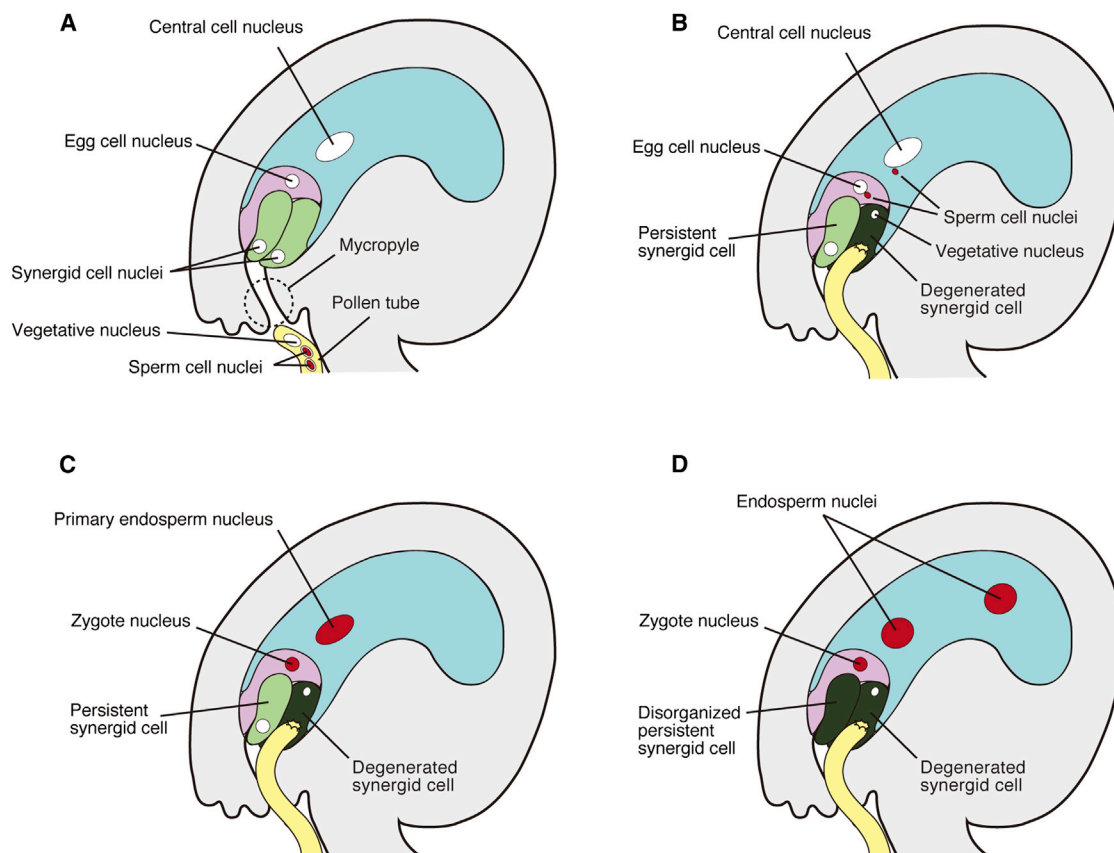


Figure 1. Diagram of Double Fertilization and Inactivation of the Synergid Cell

Diagram of double fertilization and degeneration of the synergid cells.

(A) Unfertilized ovule and pollen tube. In *A. thaliana*, AtLURE1 peptides secreted from two synergid cells attracted pollen tubes.

(B and C) Double fertilization. One of the two synergid cells received pollen tube discharge and degenerated (degenerated synergid cell). Two sperm cells fertilized either the egg cell or the central cell. Fertilization consists of plasma membrane fusion (plasmogamy, shown in B) and nuclear fusion (karyogamy, shown in C).

(D) Degeneration of the persistent synergid cell. Degeneration requires completion of double fertilization.

In the schematic, antipodal cells were omitted.

pollen tube attraction required for polytubey block. Interestingly, *Arabidopsis* is able to cancel polytubey block when the egg cell or central cell remain unfertilized, allowing the next pollen tube to recover the early fertilization failure (Beale et al., 2012; Kasahara et al., 2012; Maruyama et al., 2013).

Several *Arabidopsis* mutants display polytubey phenotype even after successful double fertilization. Ovules of the fertilization-independent seed (FIS)-class Polycomb Repressive Complex 2 (FIS-PRC2) mutant frequently receive the second pollen tube at 6 hr after arrival of the first pollen tube (Maruyama et al., 2013). FIS-PRC2 is a gene silencing complex specific to the central cell and the endosperm (Köhler et al., 2012), implying that polytubey block is activated by central cell fertilization through FIS-PRC2 pathway. Similarly, multiple pollen tube attraction as well as synergid nuclear disorganization failure were observed in *ein3 eil1* double mutant that is defective in the signaling of a gaseous hormone, ethylene (Völz et al., 2013). Although the involvement of FIS-PRC2 and ethylene in polytubey block became evident, the molecular and cellular

mechanisms by which these components control nuclear degeneration of the persistent synergid cell after successful fertilization remained elusive.

In this study, we performed live imaging of *Arabidopsis* ovules and found that degeneration of the persistent synergid cell is caused by a cell-to-cell fusion with the endosperm a few hours after fertilization. This fusion is exclusively induced by the central cell fertilization and the cytoplasm of the persistent synergid cell, including pre-secreted pollen tube attractant peptides, becomes rapidly diluted into the endosperm, suggestive of the mechanism of early cessation of pollen tube attraction. After the fusion, the persistent synergid nucleus in the endosperm exhibited disorganization synchronized with the endosperm nuclear division. We also demonstrated that the egg cell fertilization strongly activates ethylene signaling, positively controlling nuclear disorganization of the persistent synergid cell. Our data show not only a mechanism of how two female gametes independently but coordinately control polytubey block, but also a rapid and unique cell-elimination system mediated by a cell-to-cell fusion.

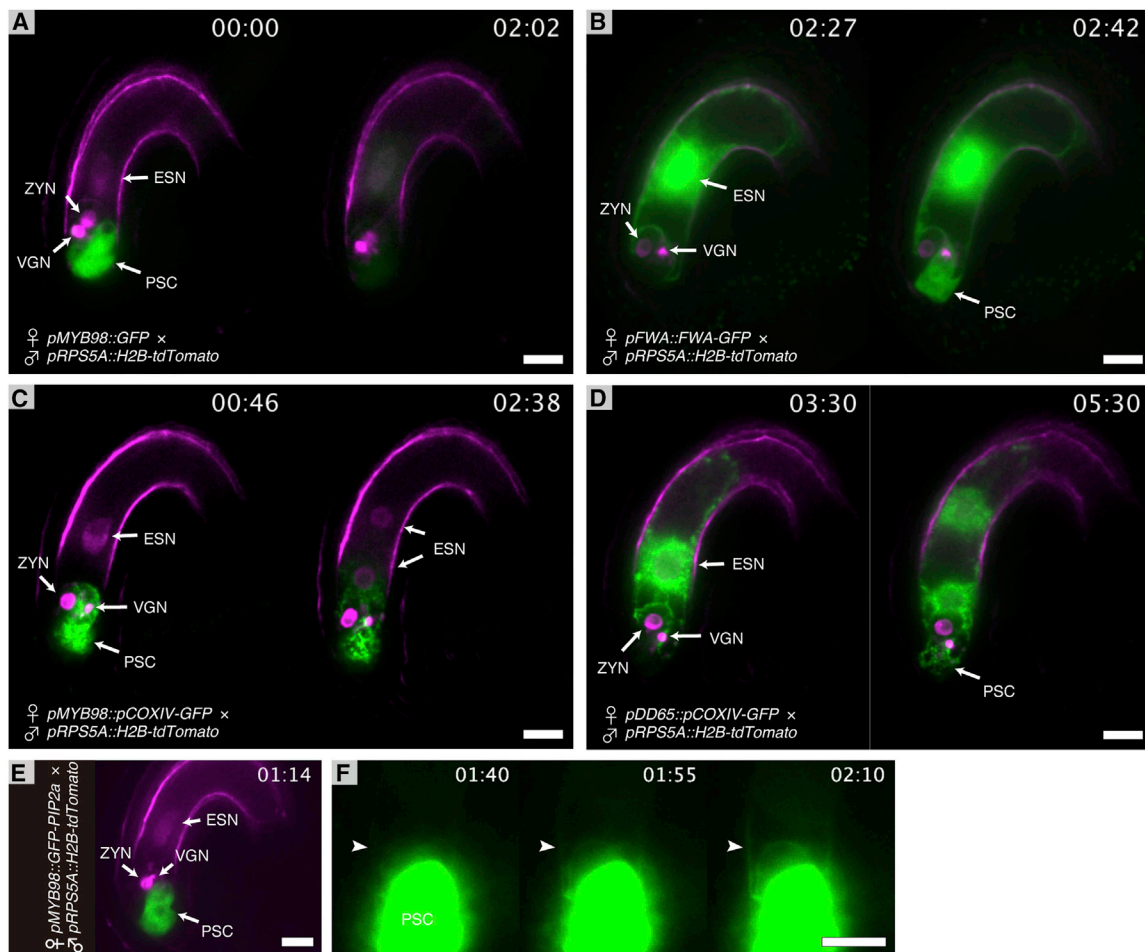


Figure 2. Cytoplasm Mixing between the Persistent Synergid Cell and Endosperm

(A and B) Dynamics of cytosol in ovules from *pMYB98::GFP*, a synergid cell marker line (A), or *pFWA::FWA-GFP*, an endosperm marker line (B) were analyzed after fertilization.

(C and D) Movement of mitochondria visualized using a *pCOXIV-GFP* fusion protein in the persistent synergid cell and endosperm. Fertilized ovules from *pMYB98::pCOXIV-GFP*, a synergid cell marker line (C), or *pDD65::pCOXIV-GFP*, an endosperm marker line (D) were analyzed.

(E) Fertilized *pMYB98::GFP-PIP2a* ovule with visualized plasma membrane and other endomembranes in the synergid cell.

(F) Time-lapse images of the magnified micropylar region in (E). Enhanced signal in the GFP channel showed migration of the marker along the endosperm silhouette (arrowheads). Time-lapse analyses in (A) to (E), a nuclear marker line *pRPS5A::H2B-tdTomato* was used as the male parent, and the numbers stamped in each frame indicate time (h: min) from the start of the observation (8 hr after pollination, HAP). ZYN, zygote nucleus; ESN, endosperm nucleus; VGN, vegetative nucleus; PSC, persistent synergid cell. Scale bars, 20 μ m.

See also [Figure S1](#) and [Movie S1](#).

RESULTS

Cell-to-Cell Fusion between the Persistent Synergid Cell and the Endosperm

To explore the degeneration mechanism of the persistent synergid cell after fertilization, pistils from a synergid cell-specific *pMYB98::GFP* marker line ([Kasahara et al., 2005](#)) were pollinated by pollens from a *pRPS5A::H2B-tdTomato* plant, a transgenic line ubiquitously expressing HISTONE 2B tagged with tdTomato ([Figure 2A](#); [Movie S1](#)) ([Adachi et al., 2011](#)). Seven hours after pollination, ovules from the pistil were cultured in liquid medium and were observed by confocal microscopy for time-lapse image analysis. Fertilized ovules were marked by the male-derived

tdTomato signal in the zygote, endosperm, and pollen vegetative nuclei ([Figure 1C](#)). Around the first endosperm nuclear division (9–11 hr after pollination [HAP]), GFP signal abruptly decreased in the persistent synergid cell and was conversely elevated in the endosperm ($n = 10$) ([Figure 2A](#); [Movie S1](#)). This GFP signal intensity shift between the persistent synergid cell and the endosperm coincided; furthermore, the GFP signal intensities in these cells became indistinguishable within 20 min after the initiation of GFP signal shift. Consistently, *pFWA::FWA-GFP*, a marker line that visualizes the endosperm ([Kinoshita et al., 2004](#)), showed an abrupt increase in GFP signal in the persistent synergid cell after fertilization ($n = 7$) ([Figure 2B](#); [Movie S1](#)). These data suggest that the persistent synergid cell and the endosperm became

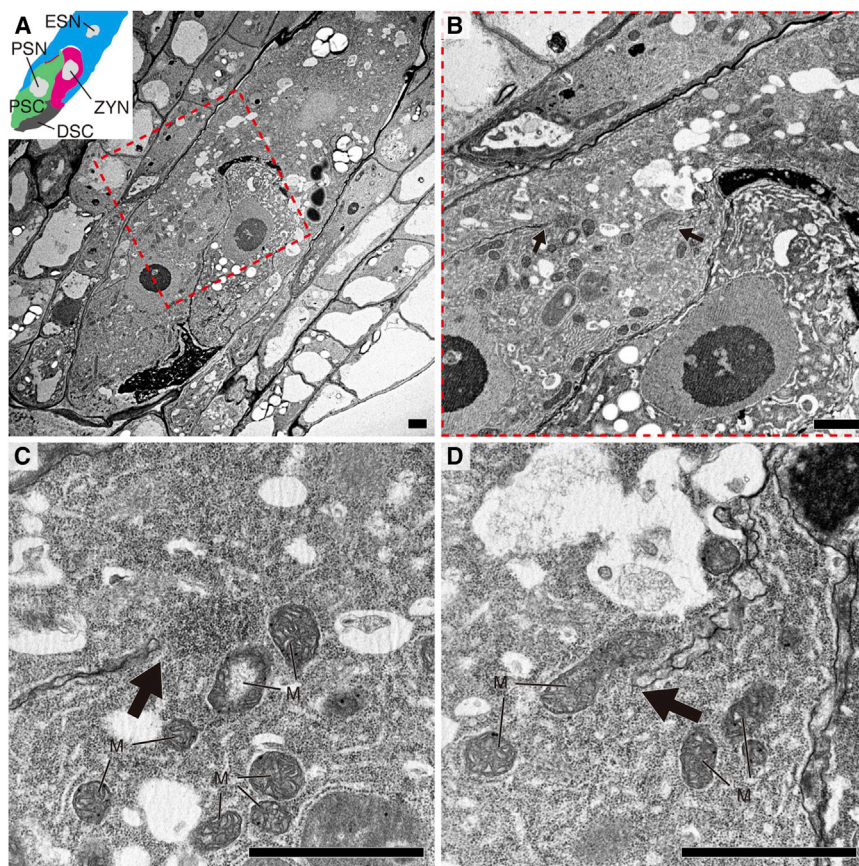


Figure 3. Electron Micrographs of Fertilized *pFWA::FWA-GFP* Ovule

(A) Image of lower magnification. Inset shows a schematic of the embryo sac components. Nuclei are indicated by light gray. The region of cell wall disintegration is shown by a red solid line.

(B) Magnification of the region of cell wall disintegration between the persistent synergid cell and the endosperm highlighted by a red dashed box in (A).

(C and D) Magnification of the disintegrated cell wall (indicated by arrows in (B)). Scale bars, 2 μ m. ZYN, zygote nucleus; ESN, endosperm nucleus; PSN, persistent synergid nucleus; PSC, persistent synergid cell; DSC, degenerated synergid cell; M, mitochondria.

See also Figure S2.

fusion between the persistent synergid cell and endosperm occurs.

To confirm the cytoplasmic continuity between the persistent synergid cell and the endosperm that happens after fertilization, the ultrastructure was analyzed by transmission electron microscopy. In the unfertilized mature ovule, the central cell and the synergid cell were separated by thin cell wall (~80 nm thickness) and their cytoplasm can be distinguished from each other with the signature of many small vacuoles in the central cell (Figures S2A–S2D). On the other hand,

fused and generated contiguous cytosol. By contrast, unfertilized *pFWA::FWA-GFP* ovules did not show any alterations in the GFP fluorescence pattern (data not shown), showing that fertilization is required for the cytosol mixture between the persistent synergid cell and the endosperm. To confirm whether the same event occurs in vivo, *pFWA::FWA-GFP* pistils were pollinated by pollens from *pRPS5A::H2B-tdTomato* and ovules collected from the pistils at 8 HAP or 12 HAP were analyzed by confocal microscopy (Figure S1). The mixture of the cytosol were observed at 12 HAP (Figures S1B–S1D) and not at 8 HAP (Figures S1C and S1D), the timing of first endosperm nuclear division and ~1 hr after fertilization, respectively. Taken together, these data indicate that the cytosol mixture is not induced immediately after fertilization, but a few hours after fertilization when the first endosperm nuclear division start.

We then monitored the diffusion of larger cell components, such as mitochondria and endoplasmic reticulum (ER), marked by fluorescence and observed the exchange of these organelles between the endosperm and persistent synergid cell (Figures 2C and 2D; Movie S1). These migrations were completed within a short period (15 min, fastest), suggestive of a rapid initiation and expansion of holes between the two cells. Furthermore, the plasma membrane marker PIP2a (Igawa et al., 2013) expressed in the synergid cell (Figure 2E) spread rapidly into the plasma membrane of the central cell after fertilization ($n = 5$) (Figure 2F; Movie S1), demonstrating that plasma membrane

electron micrographs of fertilized ovules showed the absence of a cell wall between the persistent synergid cell and endosperm (Figures 3C and 3D; width = $5.9 \pm 2.8 \mu\text{m}$, $n = 4$, mean \pm SD). Indeed, we could not find differences in the cytoplasm between the persistent synergid cell and the endosperm (Figures 3A, 3B, and S2G). Sometimes, disorganized synergid nucleus exhibiting discontinuity of its nuclear envelope was observed in the endosperm (Figures S2E–S2H), but we did not observe any defect in the cytoplasm, such as mitochondrial disorganization, indicating a selective-destruction of the persistent synergid nucleus after the cell-fusion. We consider that the unique fusion between the persistent synergid cell and the endosperm would be an important event in polytubey block and named this peculiar phenomenon synergid-endosperm fusion (SE fusion).

The Synergid-Endosperm Fusion and Pollen Tube Attractant Peptide Dynamics

Synergid cells secrete pollen tube attractant peptides, such as AtLURE1 (Takeuchi and Higashiyama, 2012). The SE fusion possibly disturbs homeostasis of the attractant responsible for the block of multiple pollen tube attraction after successful fertilization. We thus analyzed ovules expressing AtLURE1-GFP from the *AtLURE1* promoter after the ovules were fertilized by the *pRPS5A::H2B-tdTomato* pollen. As reported previously, a strong GFP signal was observed at the micropylar tip of the synergid cell (filiform apparatus) (Figure 4A) (Takeuchi and

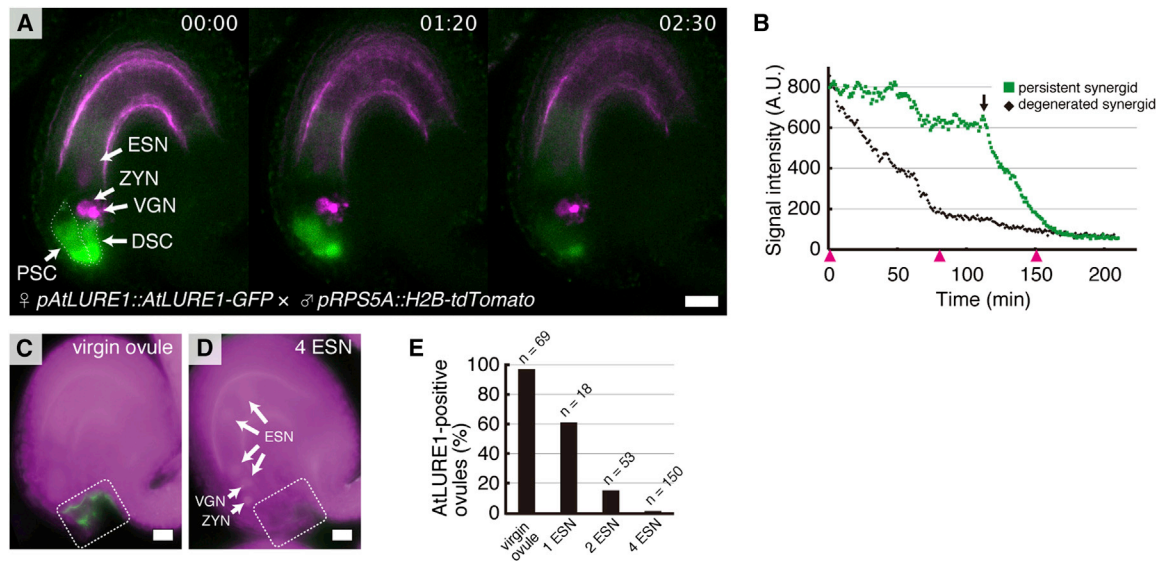


Figure 4. Rapid Dilution of Pre-Secreted Pollen Tube Attractant in the Synergid Cell Fused with Endosperm

(A) Dynamic changes of pollen tube attractant was analyzed in *pAtLURE1::AtLURE1-GFP* ovule fertilized with the *pRPS5A::H2B-tdTomato* pollen. Time stamps are as in Figure 2.

(B) Signal intensity of the two synergid cells in (A). Arrow, beginning of the SE fusion. Arrowheads show the time points of the three images in (A).

(C and D) Immunofluorescence of *AtLURE1* in virgin ovules (C) and fertilized ovules with four endosperm nuclei (D). Ovules were analyzed after pollination between wild-type plants and the *pRPS5A::H2B-tdTomato* plant. *AtLURE1* signal in the micropylar region (area in dotted line) was detected in (C), but not (D).

(E) Frequency of *AtLURE1*-positive ovules in the analysis shown in (C) and (D).

Abbreviations are as in Figure 2. Scale bars, 20 μ m.

See also Movie S2.

Higashiyama, 2012). GFP signal in the synergid cytoplasm was also detected, likely corresponding to pre-secreted *AtLURE1-GFP* (Figure 4A). In the degenerated synergid cell, the GFP signal became reduced to reach a plateau after ~80 min from the observation start (Figure 4B). By contrast, the persistent synergid cell maintained a high GFP signal and showed steep reduction at ~150 min from the observation start (Figure 4B). The half-life of the rapid reduction phase was 24 min in the persistent synergid cell, which was shorter than the degenerated synergid cell (36 min). In the filiform apparatus, decreasing of the GFP signal seemed slower in the degenerated synergid cell, which may indicate a stall of *AtLURE1* secretion after the degeneration by pollen tube reception. Taken together, these results show that *AtLURE1-GFP* signal in the persistent synergid cell decrease rapidly even compared to the degenerated synergid cell, implying a robust inactivation of pollen tube attraction in the persistent synergid cell.

We performed immunostaining against *AtLURE1* to investigate an effect of the SE fusion on polytubey block. *AtLURE1* was detected in 97% of the unfertilized ovules that exhibited no sign of pollen tube penetration (Figures 4C and 4E). The percentage of *AtLURE1*-positive ovules became <20% in ovules containing the two-nuclei endosperm and only 1% in ovules containing the four-nuclei endosperm (Figures 4D and 4E). These results are consistent with the inactivation of pollen tube attraction for polytubey block soon after fertilization. The most of the SE fusion took place during the two-nuclei endosperm stage (Figure S1), supporting the idea that rapid

dilution of *AtLURE1* by the SE fusion contributes to polytubey block.

The Synergid-Endosperm Fusion and Disorganization of the Synergid Nucleus

The synergid inactivation is marked by a loss of accumulation of nuclear protein, such as *MSI1-GFP* (Beale et al., 2012) and fluorescent proteins tagged with a nuclear localization signal (NLS) (Völz et al., 2013). To further investigate the timing of this event during polytubey block, we analyzed ovules from a double marker line carrying the *MSI1-GFP* marker (*pACT11::MSI1-GFP*) and the *pRPS5A::H2B-tdTomato* marker fertilized by wild-type male. A rapid reduction of the *MSI1-GFP* signal in the cytosol first, then in the nucleus of the persistent synergid cell (Movie S3). These suggest consecutive SE fusion and nuclear disorganization, respectively. Importantly, we observed drastic condensation of the synergid nuclear chromosomes during the loss of the *MSI1-GFP* signal in the persistent nucleus (Figures 5A and 5B; Movie S3), a hallmark of persistent synergid inactivation.

We monitored the chromosomal condensation as an indicator of the synergid nuclear disorganization in time-lapse analyses of ovules from the *pRPS5A::H2B-tdTomato* fertilized by the *pRPS5A::H2B-GFP* pollen. The endosperm nucleus exhibited increasing GFP signal, indicating de novo expression of *H2B-GFP* before the first endosperm nuclear division (Figures 5C–5H; Movie S3). The *H2B-GFP* then started to accumulate in the persistent synergid nucleus after the SE fusion, followed by an

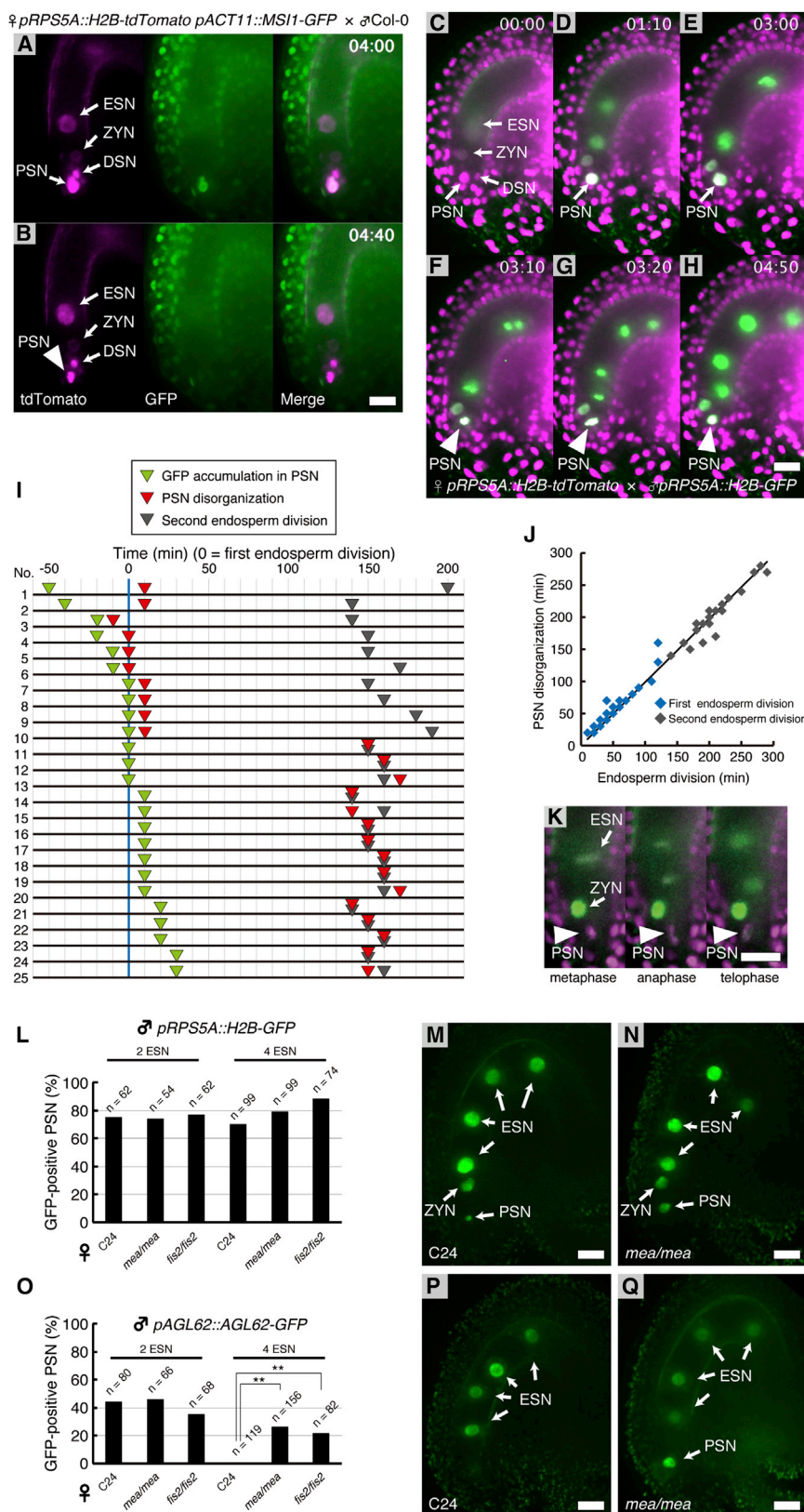


Figure 5. FIS-PRC2 Disruptions Alleviate Mitosis-Associated Disorganization of the Persistent Synergid Nucleus

(A and B) Loss of nuclear protein accumulation and chromosomal condensation in a disorganizing synergid nucleus. Morphology of nucleus and integrity of nuclear envelope were visualized by the *pRPS5A::H2B-tdTomato* marker and the *pACT11::MSI1-GFP* marker, respectively.

(C–H) Time-lapse images of a *pRPS5A::H2B-tdTomato* ovule fertilized by *pRPS5A::H2B-GFP* pollen. One-nucleus endosperm stage (C). Two-nuclei endosperm stage containing GFP-labeled persistent synergid nucleus, (D and E). Condensation of persistent synergid nucleus during metaphase (F) or anaphase (G) of the second endosperm division. Four-nuclei endosperm stage (H). Synergid nucleus changed the color from magenta (C) into white (D) and then green and white (E), indicating gradual elevation of the H2B-GFP level.

(I) A timeline chart of three cellular events. Timings of H2B-GFP accumulation in the persistent synergid nucleus (green triangles), disorganization of the persistent synergid nucleus (red triangles), and metaphase in the second endosperm division (gray triangles) are shown in each of the 25 samples. Time of metaphase in the first endosperm division was set as 0 min.

(J) Plot of the time at endosperm division and disorganization of the persistent synergid nucleus. The regression line was determined based on simple linear regression analysis ($n = 58$). Blue and gray symbols represented ovules exhibiting nuclear disorganizations during the first or the second endosperm division, respectively.

(K) Another example of an ovule displaying anaphase-associated chromosomal elongation of persistent synergid nucleus.

(L) Percentages of GFP-positive persistent synergid nucleus were analyzed at two- or four-nuclei endosperm stage in wild-type C24, *mea/mea*, and *fis2/fis2* pistils after a cross-pollination with the *pRPS5A::H2B-GFP* plants.

(M and N) Four-nuclei endosperm stage ovules from wild-type C24 (M) and the *mea/mea* mutant (N) analyzed in (L).

(O) Percentages of GFP-positive persistent synergid nucleus were analyzed at two- or four-nuclei endosperm stage in wild-type C24, *mea/mea*, and *fis2/fis2* pistils after a cross-pollination with the *pAGL62::AGL62-GFP* plants.

(P and Q) Four-nuclei endosperm stage ovules from wild-type C24 (P) and the *mea/mea* mutant (Q) analyzed in (O). Double asterisk (**), $p < 0.001$ (χ^2 test).

Disorganized persistent synergid nuclei are emphasized by arrowheads in (B), (F), (G), (H), and (K). Time stamps are as in Figure 2. ZYN, zygote nucleus; ESN, endosperm nucleus; DSN, degenerated synergid nucleus; PSN, persistent synergid nucleus. Scale bars, 20 μ m. See also Movies S3 and S4.

abrupt disorganization of the persistent synergid nucleus (Figure 5G). Interestingly, the nuclear disorganization occurred during the metaphase of either the first or the second endosperm nuclear division (Figure 5H; $R^2 = 0.98$). The persistent synergid nucleus with the H2B-GFP signal, a sign of the completion of the SE fusion, was always disorganized at the first endosperm nuclear division. However, the GFP-negative nucleus avoided disorganization during the first division (Figure 5C) and accumulated H2B-GFP during the two-nuclei endosperm stage (Figures 5D and 5E), followed by nuclear disorganization at the second endosperm nuclear division (Figure 5F). Occasionally, we also observed intermediate chromosome segregation of the persistent synergid nucleus (Figures 5G and 5K; 29%) and none had successful nuclear division (Figures 5H and 5K). These data suggested that fusion-mediated influx of mitotic signal regulated disorganization of persistent synergid nucleus.

To obtain further information for the possible involvement of mitosis, we analyzed transgenic plants containing two cell-cycle marker genes, *pHTR2::CDT1a(5G)-TagRFP* and *pCycB1;2::CycB1;2-YFP* (Movie S3) (Yin et al., 2014). The signal of the CycB1;2-YFP, an indicator of G₂/M-phase, gradually elevated in the endosperm nucleus after fertilization and disappeared during the first endosperm division. Re-accumulation of the CycB1;2-YFP signal occurred within an hour after the nuclear division, suggestive of a rapid progression of the cell cycle in the endosperm. On the other hand, the signal of the CDT1a(5G)-TagRFP, an indicator of S/G₂-phase, gradually elevated after telophase of the endosperm nuclear division. Comparing to the endosperm, the persistent synergid nucleus displayed an abrupt accumulation of CycB1;2-YFP during prophase or metaphase of the first endosperm nuclear division. In addition, the CDT1a(5G)-TagRFP signal was not observed in the persistent synergid nucleus. These results show that the persistent synergid nucleus after the SE fusion cannot establish synchronized cell-cycle status with the endosperm nuclei, likely contributing to the disorganization of the persistent synergid nucleus at endosperm nuclear division.

FIS-PRC2 Disruption Impairs Mitosis-Associated Elimination of the Persistent Synergid Nucleus

Previously, we reported a polytubey phenotype in mutants of the FIS-PRC2 components such as *mea*, *fis2*, and *fie* (Maruyama et al., 2013), indicative of defects in synergid inactivation process (e.g., SE fusion and/or nuclear disorganization). These mutants and wild-type C24 ovules were fertilized by the *prPS5A::H2B-GFP* male, and the SE fusion was monitored by an accumulation of endosperm-derived H2B-GFP in the persistent synergid nucleus. Most of wild-type ovules exhibited H2B-GFP signal in the persistent synergid nucleus as well as the zygote nucleus and the endosperm nuclei (Figures 5L and 5M). GFP-labeled persistent synergid nucleus was also observed in ~70% of ovules in the *mea* and *fis2* at the two-nuclear endosperm stage, which was comparable to the wild-type (Figures 5L–5N). Similar results were obtained at the four-nuclear endosperm stage. These data indicate that disruption of the FIS-PRC2 does not affect the SE fusion.

We then used *pAGL62::AGL62-GFP* plant as a pollen donor to analyze disorganization of the persistent synergid nucleus.

AGL62 is an endosperm-specific MADs box protein regulating endosperm proliferation (Kang et al., 2008). In wild-type ovules, AGL62-GFP signal was gradually increased in the endosperm nuclei and subsequently labeled the persistent synergid nucleus (Movie S4). Then, GFP signal disappeared from the persistent synergid nucleus during endosperm nuclear division, suggesting the loss of nuclear envelope integrity caused by nuclear disorganization (Figures 5O and 5P; Movie S4). Surprisingly, we found that significant numbers of the *mea* and *fis2* ovules displayed AGL62-GFP signal in the persistent synergid nucleus (Figures 5O and 5Q; Movie S4), indicating that the polytubey phenotype in the FIS-PRC2 mutants would be caused by a defect of endosperm-division-associated disorganization of the persistent nucleus.

Fertilization of the Central Cell Is Required for the Synergid-Endosperm Fusion

Although double fertilization is triggered by two homogeneous sperm cells, different dynamics of intracellular calcium ion between two female gametes upon fertilization implies initiation of their own activation events (Hamamura et al., 2014; Denninger et al., 2014). To identify an involvement of each female gamete for the SE fusion, ovules from the *pFWA::FWA-GFP* fertilized by the *kokopelli* mutant carrying the *prPS5A::H2B-tdTomato* marker were analyzed by time-lapse observation. The *kokopelli* mutant pollen produces aberrant sperm cells displaying a reduced fertility (Ron et al., 2010). Ovules that have received the *kokopelli* pollen tube discharge were classified into four different fertilization types determined by success or failure of fertilization in each female gamete. Double-fertilization type ovules exhibited a diffusion of sperm-derived tdTomato signal in the zygote and the endosperm (Figure 6A). Indeed, the SE fusion was observed in 87% of the double-fertilization type ovules ($n = 15$; Figure 6A) and was not in the no-fertilization type (0%, $n = 14$; Figure 6B), consistent with the result that fertilization is required for the SE fusion. The SE fusion was induced in 60% of the central-cell-fertilization type ovules ($n = 5$; Figure 6D); however, we could not observe any morphological change in the persistent synergid cell of the egg-cell-fertilization type ovules ($n = 16$; Figure 6C). These results show that fertilization of the central cell is a key signal for the induction of the SE fusion.

Fertilization of the Egg Cell Predominantly Activates the Ethylene Signaling

The ethylene signaling also controls disorganization of the persistent synergid nucleus after double fertilization (Völz et al., 2013). To examine whether the activation of the ethylene signaling is induced by a single fertilization of the egg cell or the central cell, pistils from the *pEIN3::EIN3-YFP* marker were pollinated with pollen from the *kokopelli* (*kpl*) mutant carrying the *prPS5A::H2B-tdTomato* marker gene. As reported previously, EIN3-YFP signal was detected in the persistent synergid nucleus, the zygote nucleus, and the endosperm nucleus (Figure 7B; compared to the background fluorescence of unfertilized ovule shown in Figure 7A). Although the EIN3-YFP stabilization was observed in 72% of double-fertilization type ovules ($n = 32$), none displayed the EIN3-YFP signal in no-fertilization type ovules containing tiny dots of two unfertilized sperm nuclei

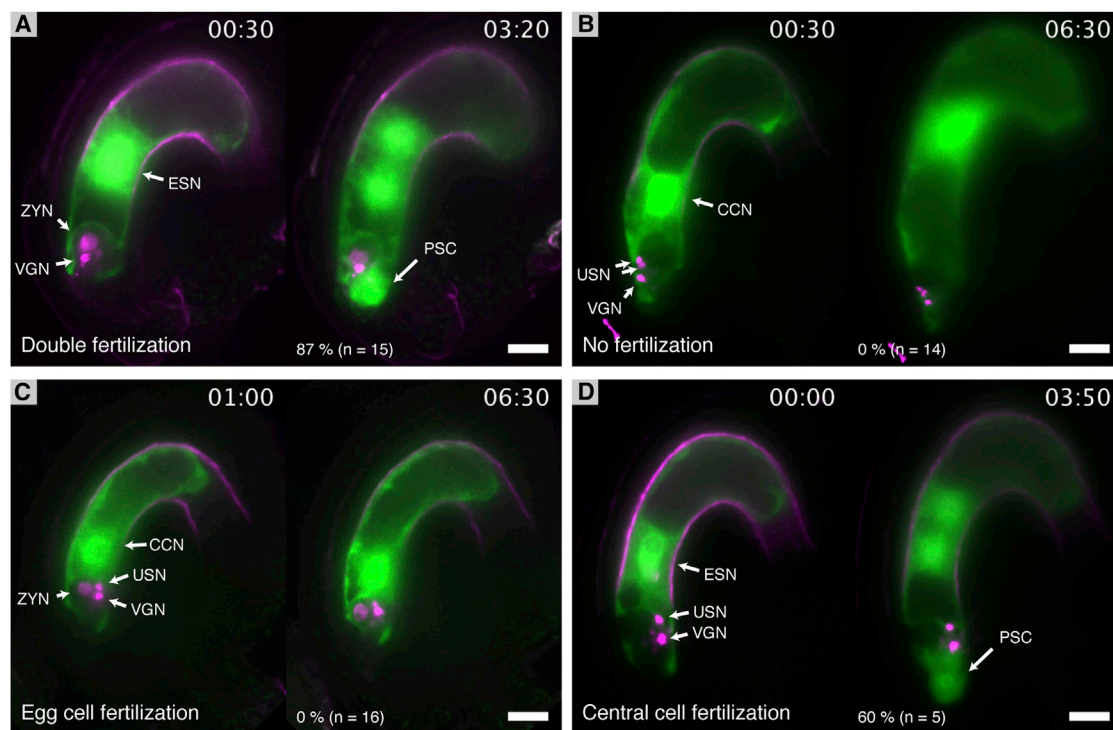


Figure 6. Synergid-Endosperm Fusion Is Induced by Fertilization of the Central Cell

(A–D) The *pFWA::FWA-GFP* plant were pollinated with pollen from the *kpl/kpl* mutant carrying the *pRPS5A::H2B-tdTomato* nuclear marker gene, and time-lapse imaging was performed to monitor the SE fusion in the ovules exhibiting different types of fertilization. Double-fertilization type (A). No-fertilization type (B). Egg-cell-fertilization type (C). Central-cell-fertilization type (D). Frequencies of SE fusion and the numbers of observed ovules are shown in the bottom of each panel. ECN, Egg cell nucleus; ZYN, zygote nucleus; CCN, Central cell nucleus; ESN, endosperm nucleus; VGN, vegetative nucleus; PSC, persistent synergid nucleus; USN, unfertilized sperm nucleus. Scale bars, 20 μ m.

(Figure 7E). We observed the EIN3-YFP signal in 64% of egg-cell-fertilization type ovules containing single unfertilized sperm cell and tdTomato-labeled zygote nucleus ($n = 73$; Figures 7C and 7E), which were comparable to the double-fertilization type. The EIN3-YFP signal was also detected in central-cell-fertilization type ovules that had tdTomato-labeled endosperm nucleus and unfertilized sperm cell (Figure 7D). However, the percentage of the YFP-positive ovule (36%) was lower than those of the double-fertilization type and egg-cell-fertilization type (Figure 7E). These results imply an ethylene-signaling-mediated polytubey block largely activated by fertilization of the egg cell.

DISCUSSION

In flowering plants, the pollen tube conveys immotile sperm cells to the female gamete cells and synergid cells play a pivotal role in attracting pollen tube toward unfertilized female gamete cells. The attraction operated by synergid cells ceases right after fertilization for polytubey block, and the elimination of the persistent synergid cell has been conjectured to be the key to this process. In this study, we found that the persistent synergid cell and the endosperm merge by a cell-to-cell fusion after fertilization. The unique plant cell-to-cell fusion, designated as the synergid-endosperm fusion (SE fusion), is a part of the polytubey blocking

system that induces rapid inactivation of the persistent synergid through a cytoplasmic dilution, followed by the selective elimination of the synergid nucleus. The egg cell and the central cell developed different pathways for the persistent synergid inactivation, highlighting a unique three-step polytubey block mechanism accomplished by double fertilization.

Discovery of the Synergid-Endosperm Fusion

One of the most prominent features of plant is the cell wall surrounding the plant cell, which has hampered the idea of cell-to-cell fusion in plants. Fusions between the two sets of gametes during double fertilization have been the only two exceptions studied extensively for more than 110 years (Strasburger, 1884; Nawashin, 1898; Guignard, 1899). Electron micrographs of *Arabidopsis* mature ovule showed very thin cell wall between the synergid cell and the central cell (Figure S2), which would be necessary for rapid digestion of the cell wall and smooth fusion of their plasma membranes. The cell wall disintegration in synergid cells was also observed in *Capsella bursa-pastoris* and barley (Schulz and Jensen, 1968; Engell, 1989; Cass and Jensen, 1970). Indeed, the absence of the boundary between the synergid cell and the endosperm in *Capsella bursa-pastoris* was reported although it was thought that the endosperm absorbed already-degenerated synergid cell (Schulz and Jensen, 1968). These results imply that the SE fusion-mediated synergid

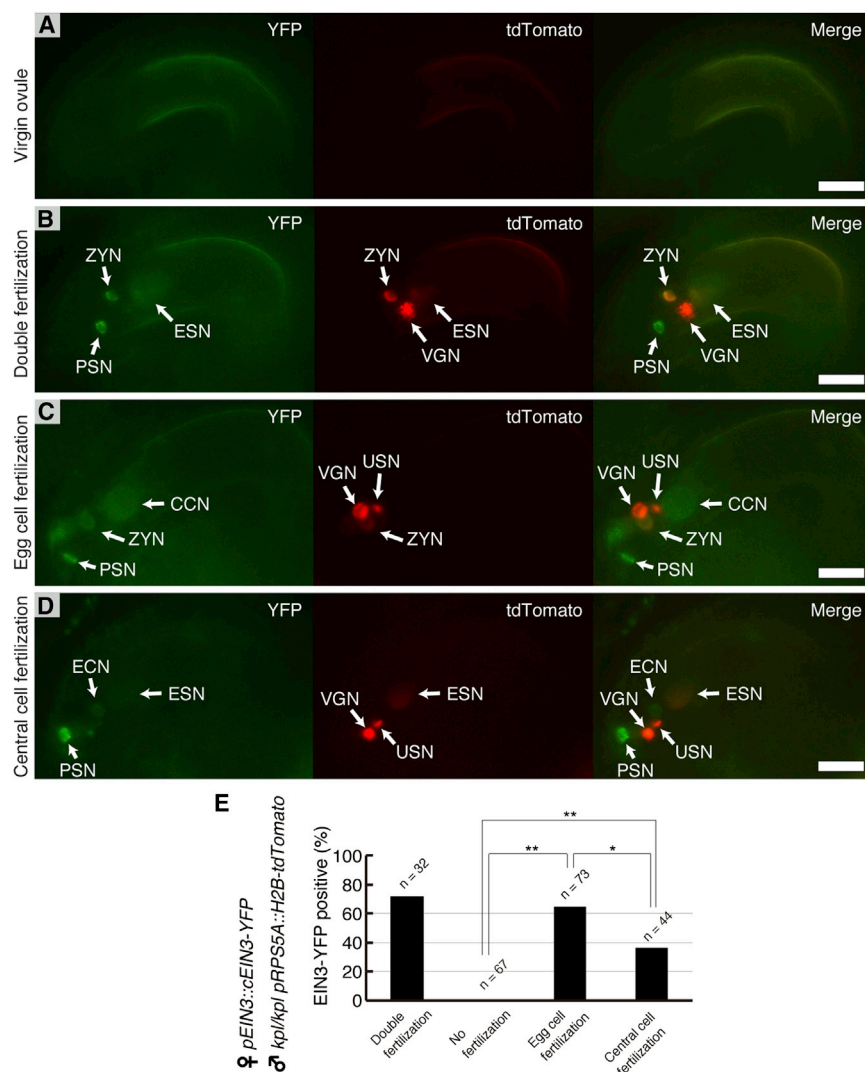


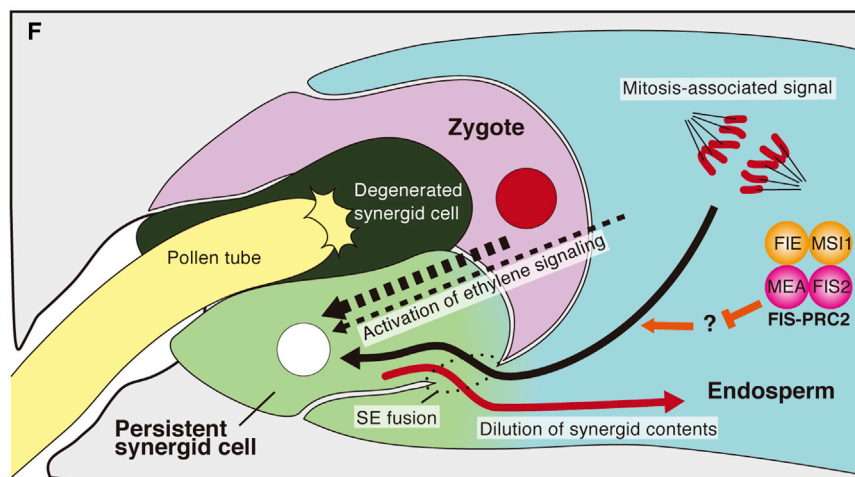
Figure 7. Fertilization of the Egg Cell Predominantly Stabilizes EIN3-YFP in the Embryo Sac

(A) Virgin ovule from the *pEIN3::cEIN3-YFP* plant. (B–D) The *pEIN3::cEIN3-YFP* plants were pollinated with pollen from the *kpl/kpl* mutant carrying the *pRPS5A::H2B-tdTomato* nuclear marker gene, and stabilizations of the EIN3-YFP were analyzed at 14 to 16 hr after pollination. These ovules exhibited different types of fertilization defects: Double-fertilization type (B); Egg-cell-fertilization type (C); and Central-cell-fertilization type (D). YFP signal predominantly accumulated in the persistent synergid nucleus.

(E) Percentages of YFP-positive ovule of each fertilization type. No-fertilization type corresponds ovules containing two condensed sperm nuclei that could not fertilize the female gametes. *p < 0.01; **p < 0.001 (χ^2 test).

(F) Schematic model of the persistent synergid inactivation. Fertilization of the central cell triggers SE fusion. The SE fusion rapidly dilutes synergid contents (red arrow) and disrupts supplying of pollen tube attractant peptides. The SE fusion also allows migration of mitosis-associated nuclear disorganization signal (black solid arrow) from the endosperm to the synergid nucleus. FIS-PRC2, an endosperm-specific polycomb gene silencing complex, would modulate the mitosis-associated nuclear elimination (orange lines and arrow). Ethylene signaling is strongly induced by a fertilization of the egg cell (thick dashed arrows), which probably coordinates the mitosis-associated signal to eliminate the persistent synergid nucleus. Ethylene signaling is less induced by the central-cell-fertilization (fine dashed arrow).

Abbreviations are as in Figure 6. Scale bars, 20 μ m.



inactivation for polytubey block is conserved in flowering plants. Nevertheless, our analyses of wild-type ovules revealed the SE fusion, the third cell-to-cell fusion event identified during normal developmental processes in *Arabidopsis*.

The mechanism of the plasma membrane fusion is largely unknown in flowering plants. A sperm cell-specific plasma membrane protein, GCS1/HAP2, is a sole factor that is thought to be directly involved in the membrane fusion during double fertilization (Mori, 2014). However, transcriptome data of the embryo sac indicate that GCS1/HAP2 is absent in the synergid cell and the central cell (Wuest et al., 2010). Thus, the plasma membrane fusion of these cells should be caused by different mechanism.

Execution of Selective Elimination of Synergid Nucleus after the SE Fusion

Nuclear disorganization has been one of the most remarkable features of the persistent synergid inactivation (Schulz and Jensen, 1968; Beale et al., 2012; Völz et al., 2013). We observed abrupt chromosomal condensation and the loss of nuclear envelope integrity after the SE fusion (Figures 5A–5J). Even though the persistent synergid nucleus shares the same cytoplasm with the endosperm nuclei, only the persistent synergid nucleus was selectively eliminated during the endosperm nuclear division. Nuclear degeneration is thought to be caused by nucleases in a programmed-cell-death of tracheary element in *Zinnia* (Ito and Fukuda, 2002) and in formation of the sieve element in *Arabidopsis* (Furuta et al., 2014). If nucleases are involved in the disorganization of the persistent synergid nucleus, there must be special mechanism(s) for the selective elimination, such as specific targeting of nucleases into the synergid nucleus or endosperm-specific resistance against the nucleases.

Alternatively, the selective nuclear elimination may be caused by premature chromosome condensation. Artificial fusion between two cells in different stages demonstrated that M phase propelled by one cell induces premature chromosome condensation, resulting in defective chromosome segregation or pulverization of chromosomes (Rao and Johnson, 1972; Szabados and Dudits, 1980). Indeed, we observed an abrupt increase in a G₂/M-marker, CycB1;2-YFP, in the disorganizing persistent synergid nucleus compared to the gradual accumulation of CycB1;2-YFP in the endosperm nucleus (Movie S3) (Yin et al., 2014). Chromosomal condensation and segregation-like behavior also support mitosis-associated synergid nucleus elimination (Figure 5K); however, determining the relevance of premature chromosome condensation should awaits precise quantification of DNA content in the synergid nucleus before and after fertilization.

Independent Pathways for the Inactivation of Persistent Synergid Cell by Two Female Gametes

By a mutant-induced single fertilization, we found that the central cell, but not the egg cell, could induce the SE fusion after fertilization (Figure 6). Indeed, ovules exhibiting single fertilization of the egg cell frequently received multiple pollen tubes (Maruyama et al., 2013), indicating an importance of the central cell in polytubey block. Besides, mutant ovules of MEA and FIS2, components of FIS-PRC2 specifically active in the central cell and the endosperm, often attracted second pollen tube even after dou-

ble fertilization (Maruyama et al., 2013). Although these mutants show normal SE fusion, significant percentages of ovules were defective in the mitosis-associated synergid nuclear disorganization (Figures 5L–5Q). These data strongly suggest that the central cell does not only induce the SE fusion, but also controls the selective nuclear disorganization, presumably by causing the persistent synergid nucleus susceptible to mitosis-associated elimination through an exposure to the factor(s) of FIS-PRC2 pathway in the endosperm.

Polytubey block is not fully activated by the central cell-fertilization, either (Maruyama et al., 2013). Interestingly, the persistent synergid nucleus in an ethylene signaling-defective *ein3 eil1* double mutant remains intact and accumulates endosperm proteins even after successful fertilization (Völz et al., 2013). In this study, we observed that fertilization of the egg cell could activate the ethylene signaling significantly compared to the central cell-single fertilization (Figure 7). Taken together, these results suggest that the egg cell fertilization activates the ethylene signaling important for the synergid nucleus disorganization. Ethylene signaling activated by the egg cell fertilization likely causes the synergid nucleus susceptible to mitosis-associated nuclear disorganization controlled partly by the FIS-PRC2 pathway. Indeed, exposure to overdose of the ethylene precursor AAC induced specific disorganization of the synergid nucleus in unfertilized ovules (Völz et al., 2013), indicating that the synergid cell nucleus is already primed to ethylene sensing for nuclear disorganization. Although it still remains unclear how ethylene signaling activated through the egg cell-fertilization and FIS-PRC2 pathway in the endosperm communicate to achieve synergid nuclear disorganization, two female gametes appear to have evolved different pathways that coordinately control selective elimination of the synergid nucleus by double fertilization. Interestingly, the persistent synergid cell in cotton is surrounded by thick cell wall and shows gradual collapse of its content (Schulz and Jensen, 1977). This may also suggest that the persistent synergid elimination through the SE fusion followed by nuclear elimination was established after the innovation of double fertilization in flowering plants.

Three-Step Polytubey Block Mechanism Mediated by Double Fertilization

In *A. thaliana*, strong polytubey block is established within a few hours (Kasahara et al., 2012). To explain the enigmatic early cessation of pollen tube attraction, we propose a three-step polytubey block system based on the analysis of AtLURE1 distribution combined with the SE fusion and dynamics change of the synergid nucleus (Figures 4 and 5). The initial step is the SE fusion caused by the central cell fertilization. After the SE fusion, pre-secreted AtLURE1 is rapidly diluted into the endosperm occupying the majority of the volume of large embryo sac, which interrupts the supply of AtLURE1. It is possible that dynamic protoplasmic streaming scrapes out synergid contents to accelerate dilution (see also fertilized *pFWA::FWA-GFP* ovule in Movie S1). Similar to the AtLURE1, other unique transcripts and proteins in the synergid cell must be diluted by the fusion (Wuest et al., 2010), by which the synergid would lose its identity rapidly. Consistently, a semi-in vitro pollen tube attraction assay for the fertilized ovule elucidated strong cessation of pollen tube

attraction prior to the disorganization of persistent synergid nucleus (Maruyama et al., 2013), indicating very early, but temporal polytubey blocking mechanism caused by the SE fusion. The second step is ethylene-signaling activation mainly by the egg cell fertilization. The persistent synergid nucleus receives this signal before/after the SE fusion, preparing for the selective nuclear elimination. The final step is mitosis-associated nuclear disorganization, which completely eliminates the source of synergid identity. The SE fusion usually occurs 9–11 hr after pollination. This is sufficiently earlier than the targeting of the second pollen tube observed in unfertilized ovules (~16 hr after pollination) (Kasahara et al., 2012).

In multicellular organisms, elimination of particular cell is important for tissue development such as the formation of digits and nervous system in animals (Milligan and Schwartz, 1997) and megasporogenesis in plants (Russell, 1979). Those examples of programmed-cell-death display characteristic degeneration processes such as cell shrinkage, chromatin condensation and organelle destruction. Although degenerative alteration was not observed in the persistent synergid except for the nucleus (Schulz and Jensen, 1968) (see also Figures 3 and S2), elimination of the persistent synergid cell has been considered as a programmed-cell-death, because it is well controlled by fertilization signals (Völz et al., 2013; Maruyama et al., 2013; Beale and Johnson, 2013). Our discoveries uncovered mysteries of independent controls for polytubey block by fertilization of two female gametes and rapid inactivation of the persistent synergid function. This study should shed lights on unique mechanisms of the cell-cell fusion and selective nuclear disorganization as well as the evolution of sexual reproduction in flowering plants.

EXPERIMENTAL PROCEDURES

Plant Materials and Growth Conditions

Col-0, Ler, and C24 were used as the wild-type plants. The *pRPS5A::H2B-tdTomato*, *pRPS5A::H2B-GFP*, *pACT11::MSI1-GFP*, *pEIN3::EIN3::YFP*, *pFWA::GFP-PIP2a* transgenic lines and double marker line of *pHTR2::CDT1a(5G)-TagRFP* and *pCycB1;2::CycB1;2-YFP* were described previously (Adachi et al., 2011; Ingouff et al., 2007; Völz et al., 2013; Igawa et al., 2013; Yin et al., 2014). The *pFWA::FWA-GFP* transgenic line was provided by T. Kinoshita (Kinoshita et al., 2004). The *pAGL62::AGL62-GFP* transgenic line was donated by G.N. Drews (Kang et al., 2008). The *kpl-2/kpl-2* mutant that is also homozygous for the *pRPS5A::H2B-tdTomato* was described previously (Maruyama et al., 2013). *mea-7/mea-7* and *fis2-6/fis2-6* seeds were kindly provided by F. Berger (Guitton et al., 2004). Plants were grown in soil at 22°C under continuous light.

Plasmids and Transgenic Plants

Constructions of plasmids and transgenic plants are described in Extended Experimental Procedures. Oligonucleotides used in this study are shown in Table S1.

Immunostaining

Wild-type pistils were pollinated by the *pRPS5A::H2B-tdTomato* plant. Immunostaining was performed 12 hr after pollination using an antibody against AtLURE1.2 protein, as described previously (Takeuchi and Higashiyama, 2012).

Transmission Electron Microscopy

Fertilized ovules were analyzed as follows. *pFWA::FWA-GFP* pistils were pollinated with the *pRPS5A::H2B-tdTomato* transgenic plant. After 9 hr, ovules were dissected from the pistils and aliened on agar pads (half-strength Mura-

shige and Skoog's medium, 5% sucrose, adjusted pH to 5.7 with 1 M KOH, 1.5% Nusieve GTG agarose). Ovules containing GFP-labeled persistent synergid cells were observed using fluorescence microscopy and selected under a dissecting microscope and subsequently fixed in 2% glutaraldehyde, 4% paraformaldehyde, and 50 mM sodium cacodylate, pH 7.4, for 3 days at 4°C. The tissue segments were washed in buffer and post-fixed for 8 hr in 2% aqueous osmium tetroxide at 4°C. The tissue was then dehydrated in a graded ethanol series, transferred into propylene oxide, infiltrated, and embedded in Quetol 651. Series of thin sections (80 nm) were stained with 2% aqueous uranyl acetate and lead citrate and examined using a JEOL JEM 1200 EX electron microscope at 80 kV.

Time-Lapse Imaging

Six HAP ovules were dissected from the pistils into half-strength Murashige and Skoog's medium (5% sucrose, adjusted pH to 5.7 with 1 M KOH). Time-lapse data was collected at 2 hr after preparation for dissection and microscopic settings. Confocal images were acquired using an inverted microscope (IX-81, Olympus) equipped with an automatically programmable XY stage (MD-XY30100T-Meta; Molecular Devices), a disk-scan confocal system (CSU-XI, Yokogawa Electric), 488 nm and 561 nm LD lasers (Sapphire, Coherent), and an EM-CCD camera (Evolve 512, Photometrics). Time-lapse images were acquired every 10 min using multiple z planes (1.5-μm intervals) and eight planes with a water-immersion objective lens (UAPo 40×W3/340; Olympus). Sequential images were acquired at 600-ms exposures for the 488-nm excitation and 200-ms exposures for the 561-nm excitation. Images were processed with Metamorph ver. 7.7.7.7.0. (Universal Imaging) to create maximum-intensity projection images.

Image Processing

Image J Ver. 1.43u (<https://www.macbiophotonics.ca/index.htm>) and QuickTime Player 7 Ver. 7.7.1 was used for movie editing of the time-lapse analyses. All other images were processed for publication using Adobe Photoshop CS ver. 8.0.1 (Adobe Systems).

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, two figures, one table, and four movies and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.03.018>.

AUTHOR CONTRIBUTIONS

T.M. and D.M. discovered the synergid-endosperm fusion. D.M. designed experiments and directed entire project. Immunostaining of the AtLURE1.2 was conducted by H.T. Stabilization of the EIN3-YFP was analyzed by R.V., R.G., and D.M. Important plant materials were constructed by T.I., D.K., M. Ueda, M. Umeda, and S.N. The other experiments were performed by D.M. The manuscript was written by D.M. and T.K. with comments from T.M., R.G., S.N., and T.H.

ACKNOWLEDGMENTS

We thank T. Kinoshita, S. McCormick, F. Berger, G.N. Drews, and T. Nakagawa for the materials, N. Iwata, T. Nishii, and T.M. Vu for assistance in preparing materials, S. Tiedemann for genetic analysis, and A. Schnittger for discussions. F. Berger also provided useful comments on the manuscript. D.M. was supported by the Global Center of Excellence program (Nagoya University). D.M. and H.T. were supported by grants 6526 and 5834 from the Japan Society for the Promotion of Science Fellowships, respectively. This work was supported in part by Japan Advanced Plant Science Network and by a grant from the Japan Science and Technology Agency (ERATO project to T.H.).

Received: January 13, 2015

Revised: February 9, 2015

Accepted: February 20, 2015

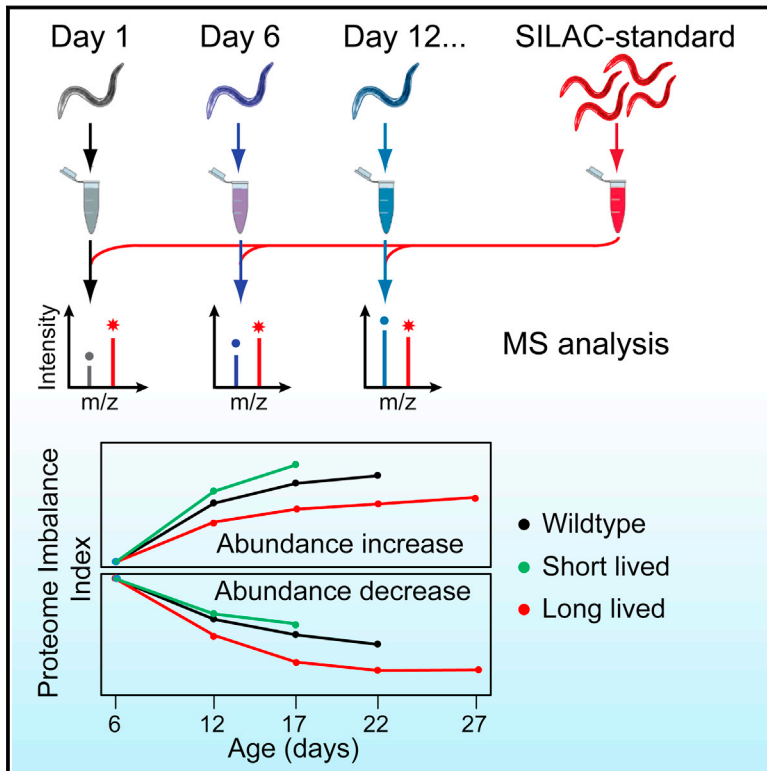
Published: April 23, 2015

REFERENCES

- Adachi, S., Minamisawa, K., Okushima, Y., Inagaki, S., Yoshiyama, K., Kondou, Y., Kaminuma, E., Kawashima, M., Toyoda, T., Matsui, M., et al. (2011). Programmed induction of endoreduplication by DNA double-strand breaks in *Arabidopsis*. *Proc. Natl. Acad. Sci. USA* 108, 10004–10009.
- Beale, K.M., and Johnson, M.A. (2013). Speed dating, rejection, and finding the perfect mate: advice from flowering plants. *Curr. Opin. Plant Biol.* 16, 590–597.
- Beale, K.M., Leydon, A.R., and Johnson, M.A. (2012). Gamete fusion is required to block multiple pollen tubes from entering an *Arabidopsis* ovule. *Curr. Biol.* 22, 1090–1094.
- Cass, D.D., and Jensen, W.A. (1970). Fertilization in barley. *Am. J. Bot.* 57, 62–70.
- Denninger, P., Bleckmann, A., Lausser, A., Vogler, F., Ott, T., Ehrhardt, D.W., Frommer, W.B., Sprunck, S., Dresselhaus, T., and Grossmann, G. (2014). Male-female communication triggers calcium signatures during fertilization in *Arabidopsis*. *Nat. Commun.* 5, 4645.
- Engell, K. (1989). Embryology of barley: Time course and analysis of controlled fertilization and early embryo formation based on serial sections. *Nord. J. Bot.* 9, 265–280.
- Furuta, K.M., Yadav, S.R., Lehesranta, S., Belevich, I., Miyashima, S., Heo, J.O., Vatén, A., Lindgren, O., De Rybel, B., Van Isterdael, G., et al. (2014). Plant development. *Arabidopsis* NAC45/86 direct sieve element morphogenesis culminating in enucleation. *Science* 345, 933–937.
- Gardner, A.J., and Evans, J.P. (2006). Mammalian membrane block to polyspermy: new insights into how mammalian eggs prevent fertilisation by multiple sperm. *Reprod. Fertil. Dev.* 18, 53–61.
- Guignard, M.L. (1899). Sur les anthérozoides et la double copulation sexuelle chez les végétaux angiospermes. *Rev. Gén. Bot.* 11, 129–135.
- Guitton, A.E., Page, D.R., Chambrier, P., Lionnet, C., Faure, J.E., Grossniklaus, U., and Berger, F. (2004). Identification of new members of Fertilisation Independent Seed Polycarb Group pathway involved in the control of seed development in *Arabidopsis thaliana*. *Development* 131, 2971–2981.
- Hamamura, Y., Nishimaki, M., Takeuchi, H., Geitmann, A., Kurihara, D., and Higashiyama, T. (2014). Live imaging of calcium spikes during double fertilization in *Arabidopsis*. *Nat. Commun.* 5, 4722.
- Igawa, T., Yanagawa, Y., Miyagishima, S.Y., and Mori, T. (2013). Analysis of gamete membrane dynamics during double fertilization of *Arabidopsis*. *J. Plant Res.* 126, 387–394.
- Ingouff, M., Hamamura, Y., Gourgues, M., Higashiyama, T., and Berger, F. (2007). Distinct dynamics of HISTONE3 variants between the two fertilization products in plants. *Curr. Biol.* 17, 1032–1037.
- Ito, J., and Fukuda, H. (2002). ZEN1 is a key enzyme in the degradation of nuclear DNA during programmed cell death of tracheary elements. *Plant Cell* 14, 3201–3211.
- Kang, I.H., Steffen, J.G., Portereiko, M.F., Lloyd, A., and Drews, G.N. (2008). The AGL62 MADS domain protein regulates cellularization during endosperm development in *Arabidopsis*. *Plant Cell* 20, 635–647.
- Kasahara, R.D., Portereiko, M.F., Sandaklie-Nikolova, L., Rabiger, D.S., and Drews, G.N. (2005). MYB98 is required for pollen tube guidance and synergid cell differentiation in *Arabidopsis*. *Plant Cell* 17, 2981–2992.
- Kasahara, R.D., Maruyama, D., Hamamura, Y., Sakakibara, T., Twell, D., and Higashiyama, T. (2012). Fertilization recovery after defective sperm cell release in *Arabidopsis*. *Curr. Biol.* 22, 1084–1089.
- Kinoshita, T., Miura, A., Choi, Y., Kinoshita, Y., Cao, X., Jacobsen, S.E., Fischer, R.L., and Kakutani, T. (2004). One-way control of FWA imprinting in *Arabidopsis* endosperm by DNA methylation. *Science* 303, 521–523.
- Köhler, C., Wolff, P., and Spillane, C. (2012). Epigenetic mechanisms underlying genomic imprinting in plants. *Annu. Rev. Plant Biol.* 63, 331–352.
- Maheshwari, P. (1950). *An Introduction to the Embryology of Angiosperms* (New York: McGraw-Hill).
- Maruyama, D., Hamamura, Y., Takeuchi, H., Susaki, D., Nishimaki, M., Kurihara, D., Kasahara, R.D., and Higashiyama, T. (2013). Independent control by each female gamete prevents the attraction of multiple pollen tubes. *Dev. Cell* 25, 317–323.
- Milligan, C.E., and Schwartz, L.M. (1997). Programmed cell death during animal development. *Br. Med. Bull.* 53, 570–590.
- Mori, T. (2014). Profiling the GCS1-based gamete fusion mechanism. In *Sexual reproduction in animals and plants*, H. Sawada, N. Inoue, and M. Iwano, eds. (Heidelberg, Germany: Springer-Verlag, GmbH), pp. 329–336.
- Nawashin, S.G. (1898). Resultate einer Revision der Befruchtungsvorgänge bei *Lilium martagon* und *Fritillaria tenella*. *Bull. Acad. Imp. Sci. St. Petersburg* 9, 377–382.
- Rao, P.N., and Johnson, R.T. (1972). Premature chromosome condensation: a mechanism for the elimination of chromosomes in virus-fused cells. *J. Cell Sci.* 10, 495–513.
- Ron, M., Alandete Saez, M., Eshed Williams, L., Fletcher, J.C., and McCormick, S. (2010). Proper regulation of a sperm-specific cis-nat-siRNA is essential for double fertilization in *Arabidopsis*. *Genes Dev.* 24, 1010–1021.
- Russell, S. (1979). Fine structure of megagametophyte development in *Zea mays*. *Can. J. Bot.* 57, 1093–1110.
- Schulz, S.R., and Jensen, W.A. (1968). Capsella embryogenesis: the synergids before and after fertilization. *Am. J. Bot.* 55, 541–552.
- Schulz, P., and Jensen, W.A. (1977). Cotton embryogenesis: the early development of the free nuclear endosperm. *Am. J. Bot.* 64, 384–394.
- Scott, R.J., Armstrong, S.J., Doughty, J., and Spielman, M. (2008). Double fertilization in *Arabidopsis thaliana* involves a polyspermy block on the egg but not the central cell. *Mol. Plant* 1, 611–619.
- Strasburger, E. (1884). Neue untersuchungen über den befruchtungsvorgang bei den phanerogamen also grundlage für eine theorie der zeugung (Jena: Gustav Fischer).
- Szabados, L., and Dudits, D. (1980). Fusion between interphase and mitotic plant protoplasts. Induction of premature chromosome condensation. *Exp. Cell Res.* 127, 442–446.
- Takeuchi, H., and Higashiyama, T. (2011). Attraction of tip-growing pollen tubes by the female gametophyte. *Curr. Opin. Plant Biol.* 14, 614–621.
- Takeuchi, H., and Higashiyama, T. (2012). A species-specific cluster of defensin-like genes encodes diffusible pollen tube attractants in *Arabidopsis*. *PLoS Biol.* 10, e1001449.
- Tsaadon, A., Eliyahu, E., Shtraizent, N., and Shalgi, R. (2006). When a sperm meets an egg: block to polyspermy. *Mol. Cell. Endocrinol.* 252, 107–114.
- Völz, R., Heydlauff, J., Ripper, D., von Lyncker, L., and Groß-Hardt, R. (2013). Ethylene signaling is required for synergid degeneration and the establishment of a pollen tube block. *Dev. Cell* 25, 310–316.
- Wuest, S.E., Vijverberg, K., Schmidt, A., Weiss, M., Gheyselinck, J., Lohr, M., Wellmer, F., Rahnenführer, J., von Mering, C., and Grossniklaus, U. (2010). *Arabidopsis* female gametophyte gene expression map reveals similarities between plant and animal gametes. *Curr. Biol.* 20, 506–512.
- Yin, K., Ueda, M., Takagi, H., Kajihara, T., Sugamata Aki, S., Nobusawa, T., Umeda-Hara, C., and Umeda, M. (2014). A dual-color marker system for *in vivo* visualization of cell cycle progression in *Arabidopsis*. *Plant J.* 80, 541–552.

Widespread Proteome Remodeling and Aggregation in Aging *C. elegans*

Graphical Abstract



Authors

Dirk M. Walther, Prasad Kasturi, ..., Matthias Mann, F. Ulrich Hartl

Correspondence

uhartl@biochem.mpg.de

In Brief

Quantitative proteomic analysis in *C. elegans* reveals widespread proteome imbalance associated with aging. Extended lifespan is correlated with increased formation of chaperone-associated aggregates, suggesting that sequestering aberrant proteins delays proteostasis decline during aging.

Highlights

- Proteome profiling reveals loss of proteome balance during *C. elegans* aging
- Proteome imbalances alter protein stoichiometries and cause proteostasis stress
- Changes in protein abundance drive extensive protein aggregation during aging
- Sequestration of surplus proteins in chaperone-associated aggregates is protective

Accession Numbers

PXD001364



Walther et al., 2015, Cell 161, 919–932
May 7, 2015 ©2015 Elsevier Inc.
<http://dx.doi.org/10.1016/j.cell.2015.03.032>

Widespread Proteome Remodeling and Aggregation in Aging *C. elegans*

Dirk M. Walther,^{1,5} Prasad Kasturi,^{2,5} Min Zheng,² Stefan Pinkert,^{2,7} Giulia Vecchi,³ Prajwal Ciryam,^{3,4} Richard I. Morimoto,⁴ Christopher M. Dobson,³ Michele Vendruscolo,³ Matthias Mann,^{1,6} and F. Ulrich Hartl^{2,6,*}

¹Department of Proteomics and Signal Transduction, Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany

²Department of Cellular Biochemistry, Max Planck Institute of Biochemistry, Am Klopferspitz 18, 82152 Martinsried, Germany

³Department of Chemistry, University of Cambridge, Cambridge CB2 1EW, UK

⁴Department of Molecular Biosciences, Rice Institute for Biomedical Research, Northwestern University, Evanston, IL 60208, USA

⁵Co-first author

⁶Co-senior author

⁷Present address: Genomics and Proteomics Core Facility, German Cancer Research Center, Im Neuenheimer Feld 280, 69120 Heidelberg, Germany

*Correspondence: uhartl@biochem.mpg.de

<http://dx.doi.org/10.1016/j.cell.2015.03.032>

SUMMARY

Aging has been associated with a progressive decline of proteostasis, but how this process affects proteome composition remains largely unexplored. Here, we profiled more than 5,000 proteins along the lifespan of the nematode *C. elegans*. We find that one-third of proteins change in abundance at least 2-fold during aging, resulting in a severe proteome imbalance. These changes are reduced in the long-lived *daf-2* mutant but are enhanced in the short-lived *daf-16* mutant. While ribosomal proteins decline and lose normal stoichiometry, proteasome complexes increase. Proteome imbalance is accompanied by widespread protein aggregation, with abundant proteins that exceed solubility contributing most to aggregate load. Notably, the properties by which proteins are selected for aggregation differ in the *daf-2* mutant, and an increased formation of aggregates associated with small heat-shock proteins is observed. We suggest that sequestering proteins into chaperone-enriched aggregates is a protective strategy to slow proteostasis decline during nematode aging.

INTRODUCTION

Protein homeostasis (proteostasis), the state in which the proteome of a living organism is in functional balance, must be tightly controlled within individual cells, tissues, and organs. Maintaining proteome balance requires a complex network of cellular factors, including the machineries of protein synthesis, folding, and degradation (Balch et al., 2008; Hartl et al., 2011), as well as neuronal signaling pathways that regulate proteostasis at the organismal level (Prahlad and Morimoto, 2009; Taylor and Dillin, 2013; van Oosten-Hawle and Morimoto, 2014). An important function of these systems is to prevent the accumulation of potentially toxic misfolded and aggregated protein species

(Knowles et al., 2014). However, as organisms age, quality control and the cellular response to unfolded protein stress become compromised (Ben-Zvi et al., 2009; Douglas and Dillin, 2010), and the defense against reactive oxygen species declines (Finkel and Holbrook, 2000). Indeed, aging is considered the principal risk factor for the onset of a number of neurodegenerative disorders associated with aggregate deposition, such as Alzheimer's, Huntington's, and Parkinson's diseases (Knowles et al., 2014). The accumulation of aberrant protein species in these pathologic states in turn places a burden on the proteostasis machinery and thus may accelerate aging by interfering with protein folding and clearance, and other key cellular processes (Balch et al., 2008; Gidalevitz et al., 2006; Hipp et al., 2014; Olzscha et al., 2011). Understanding these relationships requires systematic analyses of the changes that occur in proteome composition and balance during aging.

The nematode *C. elegans* is one of the most extensively studied model organisms in aging research, owing to its relatively short lifespan and the availability of genetic tools to identify pathways that regulate longevity. Inhibition of the insulin/insulin-like growth factor 1 signaling (IIS) pathway in strains carrying mutations in the DAF-2 receptor (or the downstream PI(3) kinase AGE-1) activates the DAF-16/FOXO transcription factor and leads to a dramatic lifespan extension (Kenyon et al., 1993; Murphy et al., 2003). Several lines of evidence suggest that the lifespan-prolonging effect of IIS reduction involves an improvement in cellular stress resistance and proteostasis capacity through upregulation of the machineries mediating protein folding and preventing the formation of toxic aggregate species (Morley et al., 2002; Cohen et al., 2009; Demontis and Perrimon, 2010). In addition to DAF-16 activation, the longevity phenotype in *daf-2* mutants requires the function of HSF-1, the transcription factor regulating the expression of multiple heat-shock proteins and molecular chaperones (Hsu et al., 2003; Morley and Morimoto, 2004). These pathways of proteostasis maintenance appear to be conserved in evolution from worms to mammals (Cohen et al., 2009; Demontis and Perrimon, 2010).

Aging and the effect of the IIS pathway have been studied in *C. elegans* by transcriptome analysis (Budovskaya et al., 2008; Golden and Melov, 2004), but only limited information exists

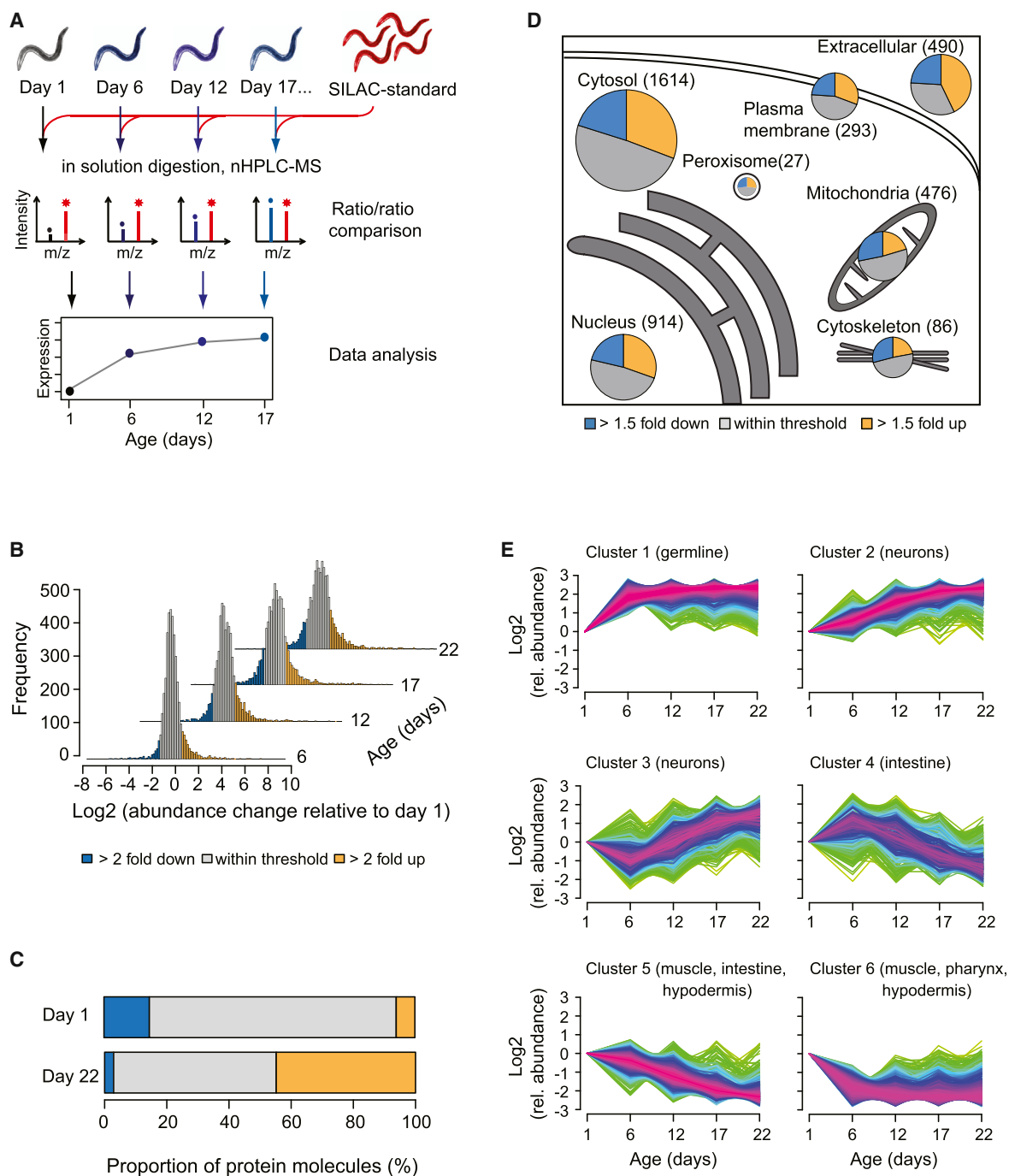


Figure 1. Proteomic Analysis of Aging in *C. elegans*

(A) Experimental design of total proteome analysis. Synchronized worm populations at different time points were lysed and mixed with a metabolically (SILAC) labeled internal protein standard. After digestion, peptides were either analyzed directly or after fractionation by isoelectric focusing, followed by nano-HPLC coupled MS.

(B) Proteome changes in WT animals 6, 12, 17, and 22 days of age relative to day 1 animals (Table S1B). The proportions of proteins that are at least 2-fold increased or decreased in abundance are marked in yellow or blue, respectively.

(C) Contribution to the total proteome of the proteins that change at least 2-fold in abundance between young (day 1) and aged (day 22) animals, as displayed in (B) and estimated by label free quantification (absolute LFQ) (Table S1B).

(D) Proteome changes in subcellular compartments. The fractions of the total proteome that increased (yellow) or decreased (blue) at least 1.5-fold in abundance in old (day 22) versus young (day 1) animals are shown. The color grey represents proteins that remained within the indicated abundance thresholds. Numbers of identified proteins are indicated. Protein subcellular localization was predicted using WoLF PSORT.

(legend continued on next page)

about changes at the proteome level (Dong et al., 2007). Here, we exploit the recent progress in mass spectrometry-based proteomics, which now enables the identification and quantification of thousands of proteins in complex mixtures (Bensimon et al., 2012; Cox and Mann, 2011). We applied stable isotope labeling with amino acids in cell culture (SILAC) (Ong et al., 2002) to profile the abundance levels of more than 5,000 different proteins at multiple time points during the lifespan of *C. elegans*. We then extended our study to short-lived and long-lived strains carrying mutations in the IIS pathway and performed a detailed analysis of age-related protein aggregation. Our data show that during aging, the proteome of the animal undergoes extensive remodeling, escaping proteostasis, and ultimately reaching a state of marked proteome imbalance. These changes are accompanied by widespread protein aggregation, with abundant proteins that exceed their solubility limit making the major contribution to aggregate load. Interestingly, the intrinsic aggregation propensity of proteins is modulated in long-lived *daf-2* mutant worms, resulting in the enhanced formation of chaperone-containing aggregates. Thus, protein aggregation may occur not just as a consequence of proteostasis decline, but may also be induced to improve proteostasis by sequestering surplus, potentially harmful protein species.

RESULTS

Extensive Proteome Remodeling during Aging

To study proteome changes in aging nematodes in depth and with high accuracy, we established a quantitative proteomics approach using SILAC (Ong et al., 2002). Near-complete incorporation of $^{13}\text{C}_6$ - $^{15}\text{N}_2$ -lysine into the proteome was achieved by feeding worms with SILAC labeled (“heavy”) *E. coli* cells (Larance et al., 2011). We used a pool of lysates prepared from labeled worms of different ages as internal standards for quantifying protein expression. These standards were added to lysates of synchronized worm populations, followed by digestion and peptide analysis by mass spectrometry (MS) (Figure 1A). Replicate analyses indicated a high degree of reproducibility between individual experiments (Figure S1A; Table S1A). We analyzed the proteomes of adult wild-type (WT) worms from 1 day up to 22 days of age, when less than 30% of the animals remain alive (L4 larval stage defined as day 0). More than 5,000 different proteins were identified and quantified at a false discovery rate of 1% (Table S1B).

Our analysis reveals a broad remodeling of the *C. elegans* proteome during aging. About one-third of the quantified proteins increased or decreased in abundance by at least 2-fold, when equal amounts of total protein were analyzed (Figure 1B; Table S1B). The proteins that increased by at least 2-fold amounted to approximately 50% of total protein in aged animals, as determined by label free absolute quantification (absolute LFQ values) (Schwanhäusser et al., 2011) (Figure 1C). Protein abundance changes were progressive until day 22 (Figures 1B and S1B;

Table S2A) and were observed in most cellular compartments (Figure 1D). Thus, proteome composition and the relative stoichiometries of proteins change dramatically during aging, presumably impeding overall proteostasis. A similar mechanism of proteostasis impairment has been suggested to occur as a result of aneuploidy (Oromendia et al., 2012; Stingle et al., 2012).

Changes in transcript levels previously observed during aging (Budovskaya et al., 2008; Golden and Melov, 2004) contribute to the changes in protein abundance observed here, but the overall correlation is only moderate ($R = 0.3$) (Figure S1C). Thus, the age-dependent accumulation of a substantial fraction of the proteome is likely to be largely due to posttranscriptional processes. Taking into consideration that microRNA (miRNA)-mediated translational repression of mRNAs is relieved during aging and stress (Ibáñez-Ventoso et al., 2006), we compared our proteome data with a published transcriptome analysis of Dicer mutant worms with defective miRNA biogenesis (Welker et al., 2007). We find that ~30% of proteins that increased more than 2-fold between day 6 and 22 (99 of 357 proteins), i.e., after the worms have reproduced, have significantly elevated transcript levels in dicer mutants, and this proportion increases to nearly 40% for the subset of proteins with a more than 4-fold abundance change (50 out of 133 proteins) (Figure S1D). Thus, miRNA-mediated translational derepression is likely to contribute to the observed increase in protein abundance.

We analyzed the proteomic changes in *C. elegans* aging in terms of various criteria, including subcellular compartments, pathways, and cell types. Among the proteins that increased more than 2-fold in aged worms (22 days) were 183 extracellular proteins (out of 490 extracellular proteins quantified) (Figure S1E; Table S2B). These included multiple transthyretin (TTR)-like factors, which increased up to 100-fold (Figure S1F), as well as all six of the vitellogenin egg storage proteins, despite egg formation having been completed before day 6. Likewise, proteins involved in DNA replication and repair processes were upregulated (Figure S1E), even though all somatic cells of adult *C. elegans* are postmitotic. These examples suggest that many changes in protein abundance during aging do not correlate with biologically relevant activities but instead reflect proteome dysregulation. Among the proteins that declined during aging are nucleolar ribosome biogenesis factors, various peroxisomal enzymes, and proteins involved in lipid glycosylation (Figure S1E; Table S2C). The levels of many mitochondrial proteins also decreased (Figure 1D). For example, subunits of respiratory chain complex I declined gradually by up to 50% during the lifespan (Figure S1G), which may result in the production of reactive oxygen species.

To discern cell-type specific patterns of change, we grouped proteins into clusters using the fuzzy c-means method (Kumar and Futschik, 2007) and analyzed these by tissue-specific expression scores (Chikina et al., 2009) (Figure 1E). We find that age-dependent changes in proteome composition affect

(E) Clustering of time course expression patterns in WT animals using the fuzzy c-means algorithm (Kumar and Futschik, 2007). Significantly enriched tissues as determined by Wilcoxon rank sum test at 2% false discovery rate against predicted expression scores (Chikina et al., 2009) are indicated for each cluster. Warm (red) and cold (blue) colors indicate low and high deviation from the consensus profile, respectively. See also Figure S1 and Tables S1 and S2.

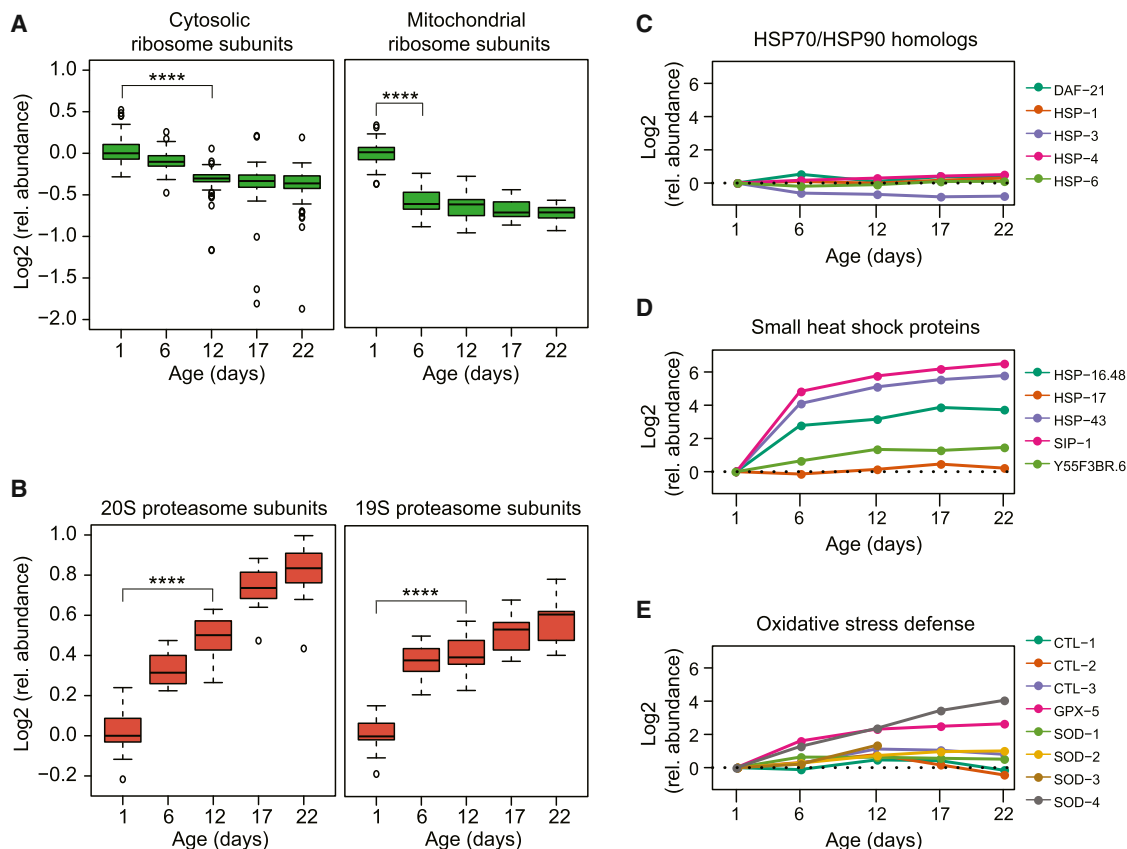


Figure 2. Abundance Changes in Specific Components of the Proteostasis Network

(A) Abundance changes of ribosomal proteins during the lifespan of *C. elegans*. There were 70 different cytosolic (left) and 34 mitochondrial ribosomal proteins (right) that were quantified (see Table S3). Log2 values of fold-changes are shown in boxplot representation. Solid horizontal lines indicate the median values, whisker caps indicate 10th and 90th percentiles, and circles indicate outliers. ****p < 4.35 × 10⁻¹³ for cytosolic ribosomal proteins and 1.17 × 10⁻¹⁰ for mitochondrial ribosomal proteins from Wilcoxon signed rank test. Only proteins quantified at both time points tested were considered.

(B) Abundance changes of proteasome subunits during lifespan. All 14 subunits of the 20S and 17 subunits of the 19S proteasome were quantified. Only subunits quantified in at least two time points are displayed. ****p < 1.23 × 10⁻⁴ for 20S subunits and ****p < 1.53 × 10⁻⁵ for 19S subunits from Wilcoxon signed rank test. (C–E) Abundance profiles of proteostasis network (PN) components along the lifespan of WT animals. Log2 relative changes in abundance are shown for HSP70 and HSP90 homologs (C), small HSPs (D), and proteins involved in oxidative stress defense (E). Only components quantified at day 1 and at least three consecutive time points are displayed.

See also Figure S2 and Tables S1 and S3.

a range of tissues. For example, proteins that are predominantly expressed in the germline strongly increase during the first 6 days of adulthood (cluster 1), when the animals reproduce, but surprisingly retain constant levels later in life. Proteins enriched in neuronal cells either increase in abundance throughout the lifespan or after day 6 (clusters 2 and 3). In contrast, the levels of many proteins enriched in intestine, muscle, and hypodermis decline (clusters 4–6), consistent with an age-related deterioration of these tissues.

Age-Related Changes in Proteostasis Network Components

Approximately 440 proteostasis network components involved in protein synthesis, folding, and degradation were quantified throughout the nematode lifespan (Figure S2A; Table S3). A ~25% reduction in the median level of cytosolic ribosomal

proteins occurred between day 1 and day 12 (Figure 2A, left). This reduction correlates with a decrease in the transcript level of ribosome proteins (Golden and Melov, 2004) and an overall age-associated reduction in polysomes (Kirstein-Miles et al., 2013). A similar decrease was observed for mitochondrial ribosomes between day 1 and day 6 (Figure 2A, right). Interestingly, aged animals displayed a pronounced imbalance in the relative subunit stoichiometry of cytosolic, but not mitochondrial, ribosomes, with several subunits decreasing more than 60% below median subunit levels (Figure 2A, left).

Next, we employed SILAC to estimate protein synthesis in aging *C. elegans*. Pulse labeling of worms with heavy bacteria as the food source showed a sharp reduction in the incorporation of labeled amino acids into protein between day 1 and day 4 of adulthood (Figure S2B; Table S1C). This effect was not caused by reduced food uptake, as *eat-2* mutant animals, deficient in

pharyngeal pumping, showed protein labeling equivalent to WT controls, despite their reduced food uptake (data not shown). The reduction in protein synthesis between day 1 and 4 is greater than the decrease in ribosomal levels (Figures 2A and S2B) and probably reflects the reduction in growth of the animals.

In contrast to the effect on ribosomes, we observed an age-dependent increase in 20S and 19S proteasomal subunits (~2-fold at day 22 for 20S subunits) (Figure 2B), correlating with an increase in proteasome activity measured in worm lysates *in vitro* (Figure S2C). Many E3 ubiquitin ligases and other components of the ubiquitin proteasome system (UPS) also increased moderately (Table S3B), while there was no systematic change in the components of autophagy (Figure S2D).

Age-dependent changes in the levels of abundant cytosolic chaperones of the HSP70 and HSP90 (DAF-21) families (Figure 2C) as well as their DnaJ (DNJ/HSP40) and tetratricopeptide repeat (TPR) co-factors were limited (Figures S2E and S2F). Similarly, the subunits of the TRiC/CCT chaperonin remained unchanged (Figure S2G). In contrast, multiple small HSPs, chaperones that function by buffering aggregation, increased dramatically (~13–90-fold), mainly between day 1 and day 6 (Figure 2D). Several of these proteins are under regulation by DAF-16 and HSF-1 (Hsu et al., 2003).

Several components mediating the defense against oxidative stress, including glutathione peroxidase isoform GPX-5 and superoxide dismutases (SOD), increased during aging (up to 12-fold) (Figure 2E; Table S3B). While changes in mitochondrial proteostasis components were generally moderate (Figure S2H; Table S3B), we observed diverse alterations in the proteostasis network of the ER during the nematode lifespan (Figure S2I). For example, protein disulfide isomerases (PDI-2 and C14B9.2), the chaperone calreticulin (CRT-1), as well as the HSP70 homolog HSP-3 decreased ~2-fold, and the pro-collagen modifying enzymes lysyl hydroxylase (LET-268) and prolyl-4-hydroxylase α (DPY-18 and PHY-2) decreased ~3–10-fold. These findings suggest an age-dependent decline in ER quality control and collagen synthesis capacity.

In summary, the levels and activities of two main branches of proteostasis control, protein synthesis and degradation, change in opposite directions during aging. The decrease in ribosomal subunit proteins is accompanied by a dysregulation of cytosolic ribosome assembly, while the increase in proteasome subunits is likely to reflect an attempt at removing surplus or damaged proteins. Other notable changes in the proteostasis system include an increase in the abundance of small HSP chaperones and of components involved in the defense against oxidative stress, as well as a decline in ER protein quality control machinery.

Proteome Changes in Long-Lived and Short-Lived Mutant Strains

To understand in more detail the relationship between the observed proteome changes during the lifespan and the aging process, we next analyzed the proteomes of long-lived *daf-2* (*e1370*), short-lived *daf-16* (*mu86*), and *hsf-1* (*sy441*) mutant worms. The increase in levels of specific proteins observed during aging of WT animals was considerably less pronounced in *daf-2* mutant animals and enhanced in *daf-16* mutant animals

(Figure S3A, left), indicating that the long-lived *daf-2* mutant strain is more effective in controlling the accumulation of surplus proteins. The extent to which proteins decreased in abundance during aging was also greater in *daf-2* mutant worms (Figure S3A, right).

The changes in components of the proteostasis network observed in the mutant strains occurred again predominantly in the protein synthesis and degradation pathways, but at different rates compared to WT. The upregulation of proteasomal subunits commenced earlier during the lifespan of the *daf-2* mutant and was more pronounced than in the WT worms (Figures 3A and 3B); such upregulation was instead less prominent in the short-lived *daf-16* and *hsf-1* mutant strains (Figures 3C and 3D). These results are consistent with the DAF-16 dependent regulation of some proteasome subunits, including RPN6, which is required for 26S proteasome assembly (Vilchez et al., 2012). The decrease in ribosomal proteins occurred at a similar rate in *daf-2* mutant worms as in WT (Figure 3A), but was strongly enhanced in *daf-16* mutant worms (Figure 3C), suggesting that DAF-16 is involved in ribosome maintenance.

Components involved in the oxidative stress response showed marked differences in levels between WT and *daf-2* mutant animals. For example, cytosolic (CTL-1 and CTL-3) and peroxisomal (CTL-2) catalases were 4–8-fold higher in the *daf-2* mutant than in WT worms throughout their lifespans (Figure S3B). SOD-1 (cytoplasmic) and SOD-2 (mitochondrial) were elevated 2-fold compared to WT and short-lived mutant animals (Figure S3C), consistent with their DAF-16-dependent transcriptional regulation (McElwee et al., 2003; Murphy et al., 2003). Among the small HSPs, SIP-1 was already more abundant in young *daf-2* mutant worms (day 1) and HSP-16.48 was markedly elevated in *hsf-1* mutant animals (Figure S3D).

The earlier and more pronounced increase in proteasome abundance in *daf-2* mutant animals may improve the capacity of the organism for the clearance of surplus proteins that accumulate during aging. The elevated levels of catalases and SOD may provide improved defense against oxidative damage.

Age-Dependent Protein Aggregation and Its Relation to Protein Abundance

Declining proteostasis capacity is thought to result in the accumulation of protein aggregates, consistent with recent reports of age-dependent aggregate formation in *C. elegans* (David et al., 2010; Reis-Rodrigues et al., 2012). To analyze this process systematically, we developed a sensitive method for the quantification of aggregated proteins (see Experimental Procedures) and validated it in animals expressing muscle specific FlucDM-GFP, a conformationally unstable mutant of firefly luciferase fused to GFP (Gupta et al., 2011) (Figures S4A and S4B). We isolated insoluble proteins from total lysates of WT animals by centrifugation and performed MS analysis using lysate from labeled worms for quantification (Figure S4C). About 90% of the proteins that were quantified in three out of four experiments (975 of 1,083 proteins) accumulated significantly in the insoluble fraction of day 12 animals relative to day 1 (Table S1D). Age-dependent aggregation was most pronounced between day 6 and day 12 (Figure 4A), i.e., after the hermaphrodite animals ceased to lay eggs. Proteins with predicted transmembrane

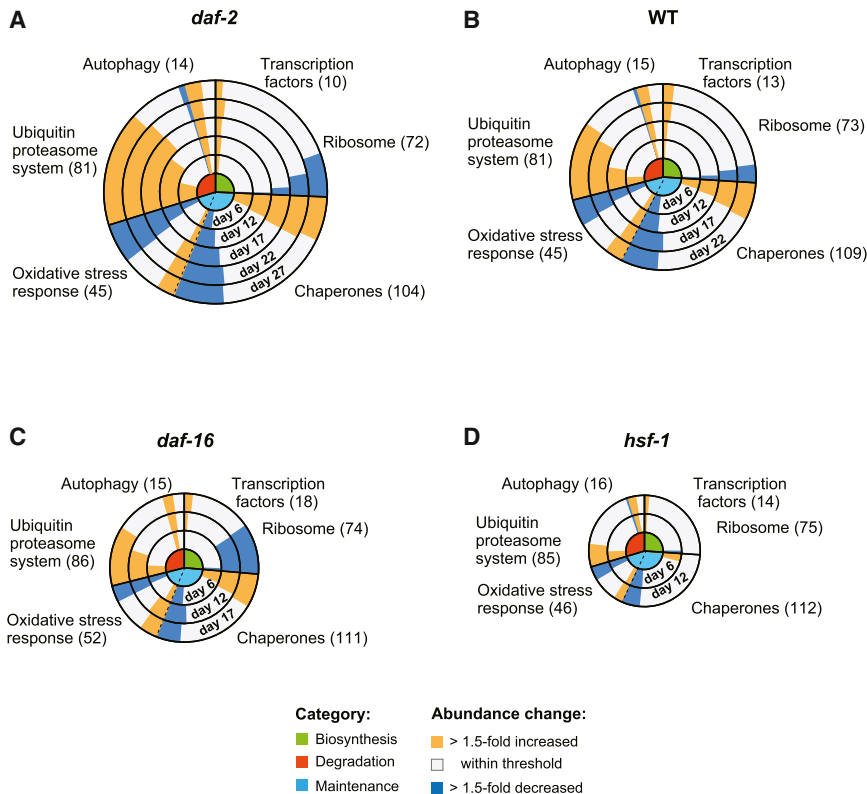


Figure 3. Remodeling of the Proteostasis Network during Aging

(A–D) Abundance changes in components of the PN (see Figure S2A) during aging in *daf-2* (A), WT (B), *daf-16* (C), and *hsf-1* (D) mutant worms. Concentric circles represent increasing age in days from center to periphery. Circle size corresponds with lifespan. Functional categories of components are indicated in the center: green, biosynthesis; red, degradation; and light blue, conformational maintenance (see Figure S2A). Abundance changes of components within these categories relative to day 1 of each strain (yellow, >1.5-fold up and blue, >1.5-fold down) are indicated as bars, with the length of the bar representing the number of proteins undergoing change. The total numbers of proteins quantified in the respective categories are indicated. See also Figure S3.

segments were not enriched in the insoluble fraction (Figure S5A), indicating that lysis was efficient.

To measure the aggregation propensities of proteins during aging, we quantified the insoluble amount of each protein as a fraction of its total amount in aged WT worms (day 12) (Figure S4D; Table S1E). The aggregation propensities of >2,100 analyzed proteins varied by more than two orders of magnitude (Figure 4B), with the median insoluble fraction per individual protein amounting to ~9%.

Previous studies reported a negative correlation between computationally predicted aggregation propensities and protein abundance (Tartaglia et al., 2009). To investigate this dependency at the proteome scale, we grouped proteins according to their aggregation propensities measured at day 12 and estimated the total abundance of each protein in the whole cell lysate by absolute LFQ (Figure 4C). The most abundant proteins were 10-times more soluble than the least abundant proteins. An analysis of the physicochemical properties of the abundant proteins based on their amino acid sequences revealed that they were more hydrophobic (Figure S5B) and more structured (data not shown) than the less abundant ones. These results suggest that abundant proteins increase their solubility, at least in part, by stabilizing their native states through formation of a more extensive hydrophobic core. Indeed, a calculation of the aggregation propensities (Z scores) (Tartaglia et al., 2008; Sormanni et al., 2015a) (see Extended Experimental Procedures) predicts that the more abundant proteins, if correctly folded, are also more soluble (Figure S5C). This conclusion is consistent

with the idea that the solubility of proteins follows their abundance (Tartaglia et al., 2009).

We found, however, that despite of their lower intrinsic aggregation propensities, the most abundant proteins contribute most to the total aggregate load. A strong correlation ($R = 0.75$) was observed between the abundance of specific proteins in the aggregate fraction

and their level in the corresponding whole cell lysate (Figure 4D). Apparently, the high solubility of abundant proteins is insufficient to protect them from age-dependent aggregation, as eventually these proteins exceed their critical concentrations, a phenomenon referred to as “supersaturation” (Ciryam et al., 2013). Notably, we also observed a medium correlation ($R = 0.43$) between the age-dependent change in the total abundance of proteins and their increase in the aggregate fraction (Figure 4E), and this correlation became stronger as aging progressed (data not shown). Thus, proteome remodeling during aging likely drives the aggregation of numerous proteins.

We further investigated whether aggregation also correlates with function. Gene ontology analysis showed that proteins with a relatively high aggregation propensity in aged animals are enriched in the nucleus, whereas abundant glycolytic enzymes and mitochondrial proteins tend to be highly soluble (Figure S5D; Table S4A). Interestingly, all identified small HSPs, but not other chaperones, were highly insoluble at day 12 (Figure 4F), with a high rate of accumulation in the aggregate fraction during aging (Figure S5E). The recruitment of these chaperones into the insoluble fraction may reflect an attempt of the organism to sequester protein aggregates.

Protein Aggregation in Long-Lived and Short-Lived Mutant Strains

Is the age-dependent formation of insoluble aggregates merely a reflection of declining proteostasis capacity, or is it a means to improve proteostasis by sequestering surplus proteins?

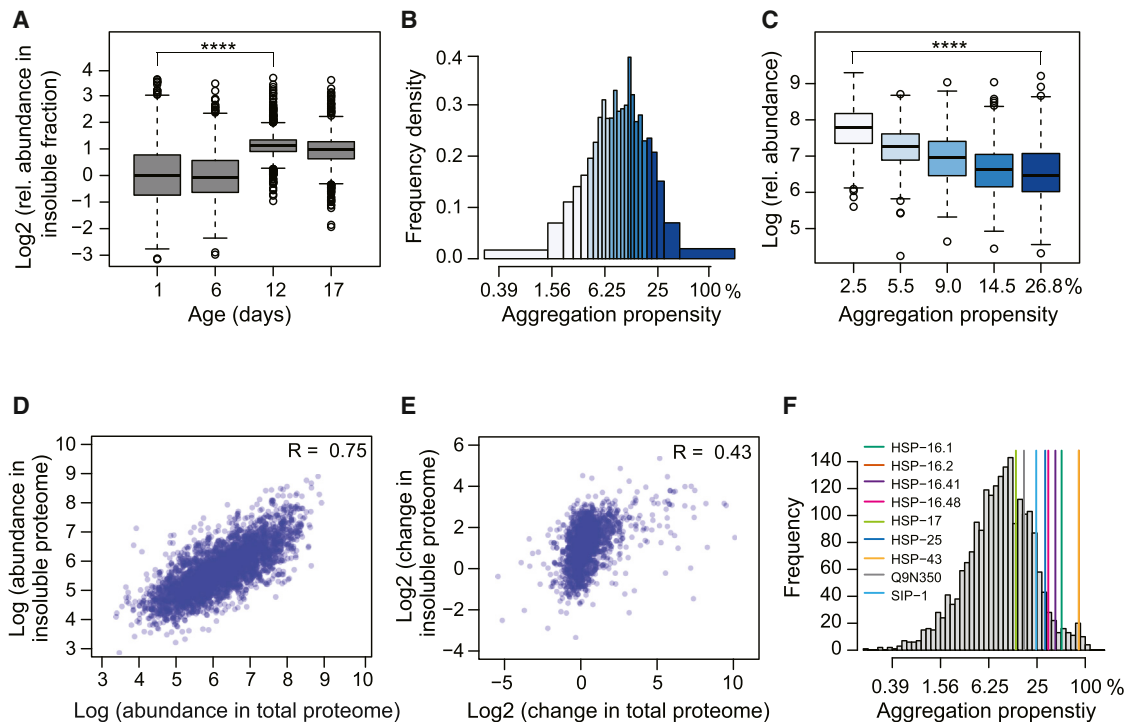


Figure 4. Proteome-wide Analysis of Protein Aggregation during Aging

(A) Relative abundance of proteins in the insoluble fraction of WT animals during aging determined by SILAC quantification (see Figure S4C; Table S1D). At least 1,355 proteins were quantified at the different time points (~3,228 different proteins in total). **** $p < 2.2 \times 10^{-16}$ from Wilcoxon signed rank test.

(B) Distribution of aggregation propensities of proteins (insoluble protein as fraction of total protein) in WT animals at day 12 (median from three independent experiments; Table S1E). Whole worm lysates and insoluble fractions were quantified against the same SILAC standard and ratios were calculated for each protein in % of total (see Figure S4D).

(C) Relationship between aggregation propensity and total protein abundance. Proteins were divided into quantiles based on their measured aggregation propensities (median values are indicated in %). LFQ was used to estimate total protein abundance (displayed as relative abundance values). **** $p < 2.2 \times 10^{-16}$ from Wilcoxon rank sum test.

(D) Protein abundance in the insoluble fraction is positively correlated with abundance in the total proteome (absolute LFQ values). Data for WT animals at day 12 are shown. The Pearson correlation coefficient R is indicated.

(E) Positive correlation between age-related protein abundance changes in the insoluble fraction and abundance changes for the same proteins in the total proteome. Abundance differences measured by SILAC between aged (day 12) and young (day 1) WT animals are plotted. The Pearson correlation coefficient R is indicated.

(F) Aggregation propensities of small HSP family members relative to the aggregation propensities of all quantified proteins in the proteome of day 12 WT animals. See also Figures S4 and S5 and Tables S1 and S4.

Consistent with the former possibility are findings that aggregation-prone model proteins increasingly aggregate in proteostasis-compromised *hsf-1* mutant strains (Ben-Zvi et al., 2009). Indeed, compared to WT animals, the short-lived *hsf-1* mutant worms accumulated more insoluble proteins and aggregation occurred earlier during aging (between day 1 and day 6) (Figures 5A and S6A). However, in support of a beneficial role for aggregation, we found that the long-lived *daf-2* mutant worms also accumulated more insoluble proteins than age-matched WT animals (Figures 5A, S6A, and S6B). This effect was not observed in *daf-16* mutant animals (Figures 5A and S6A), suggesting that age-dependent aggregation is (at least in part) an active process under regulation by DAF-16. The increased aggregation in *daf-2* mutant animals comprised preferentially cytosolic proteins (Figure S6C; Table S4B) and initiated between day 6 and day 12 as in WT (Figure S6A), i.e., when the long-lived mutant worms are still youthful.

While there was a large overlap between the proteins identified in the insoluble fractions, the extent to which specific proteins aggregated varied greatly in a strain specific manner. Interestingly, the proteins that showed increased aggregation in the *daf-2* mutant over WT are not generally more abundant at the total proteome level (Figure 5B), indicating that abundance in this case is not the main driver of aggregation. Similar findings were made in the *hsf-1* mutant (Figure 5C). On the other hand, proteins that aggregated less in the *daf-2* strain than in WT are also generally less abundant (Figure 5B), which would allow these proteins to maintain solubility.

Next, we compared the physico-chemical properties of the insoluble proteins. Strikingly, the proteins that aggregate most in the *daf-2* mutant animals are predicted to have significantly lower aggregation-propensity Z scores, are more charged, display more structural disorder (coil average) (Sormanni et al., 2015b), and are less hydrophobic compared to the proteins

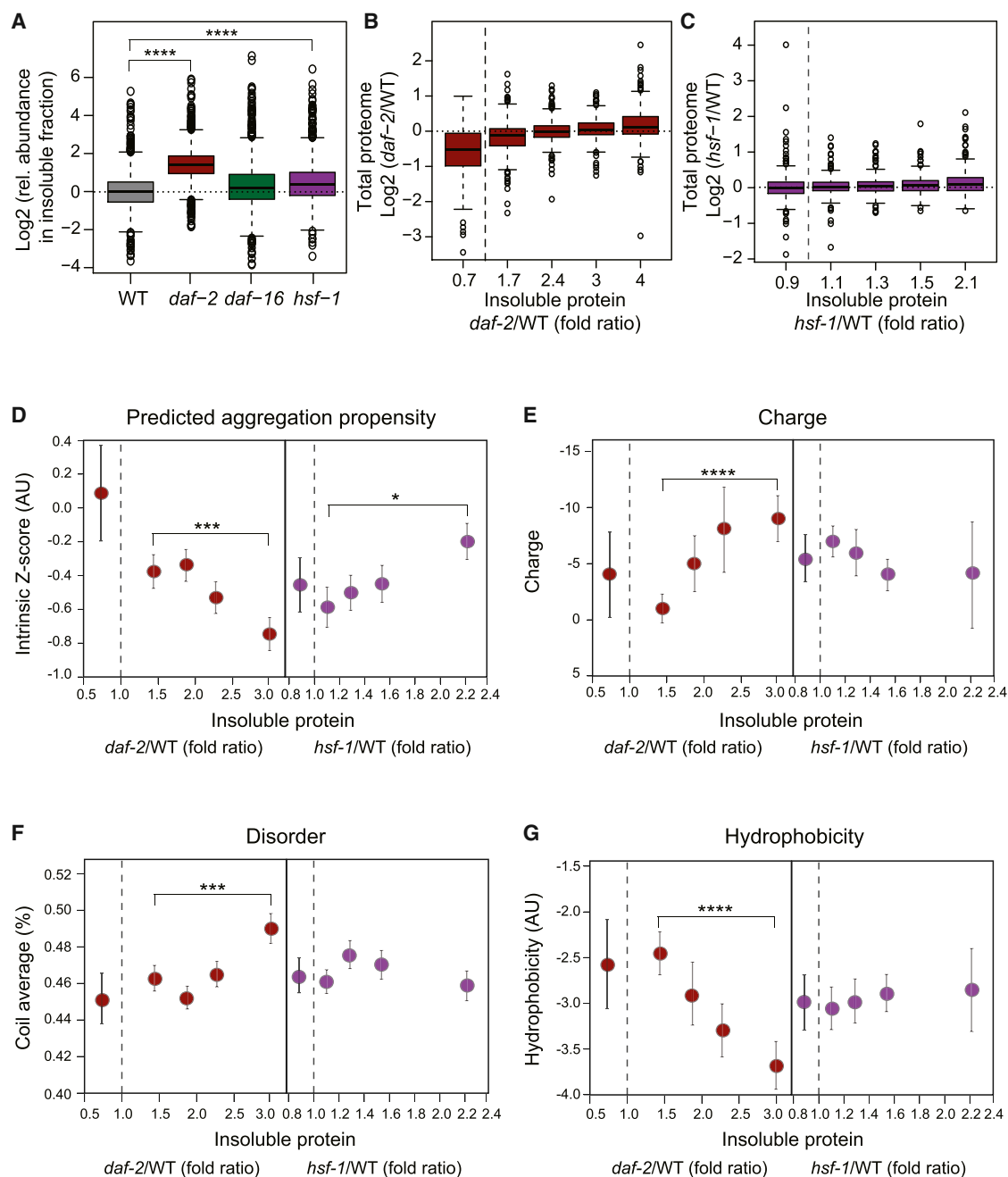


Figure 5. Protein Aggregation in Lifespan Mutant Worms during Aging

(A) Increased aggregate load in *daf-2* mutant animals compared to WT, *daf-16*, and *hsf-1* mutant animals at day 12. Relative abundance values of proteins in the insoluble fraction were determined by SILAC quantification. There were 1,367, 1,988, 1,449, and 1,485 proteins that were quantified in WT, *daf-2*, *daf-16*, and *hsf-1* mutant animals, respectively (one representative out of four independent experiments is displayed; [Table S1F](#)). **** $p < 2.2 \times 10^{-16}$ from Wilcoxon signed rank test. (B and C) Quantified abundance of proteins in the insoluble fraction of *daf-2* (352–354 proteins per quantile) (B) and *hsf-1* mutant (292 proteins per quantile) (C) relative to WT animals at day 12 plotted against differences in total protein abundance values. Quantile median values are indicated on the x axis. Proteins that aggregated less in the mutant strains than in the WT have been grouped separately (91 proteins in *daf-2* and 259 in *hsf-1* mutant).

(D–G) Physico-chemical properties of proteins enriched in the insoluble fractions of *daf-2* and *hsf-1* mutant relative to WT animals at day 12.

(D) Aggregation propensity scores (intrinsic Z scores, see [Extended Experimental Procedures](#)). *** $p < 1.4 \times 10^{-4}$ and * $p < 0.016$, Wilcoxon rank sum test.

(E) Net charge. **** $p < 4.9 \times 10^{-11}$.

(F) Coil content. *** $p < 1.1 \times 10^{-4}$.

(G) Overall hydrophobicity. **** $p < 2.9 \times 10^{-7}$. Quantile median values are indicated on both axes and standard errors are reported on the y axis.

See also [Figure S6](#) and [Table S4](#).

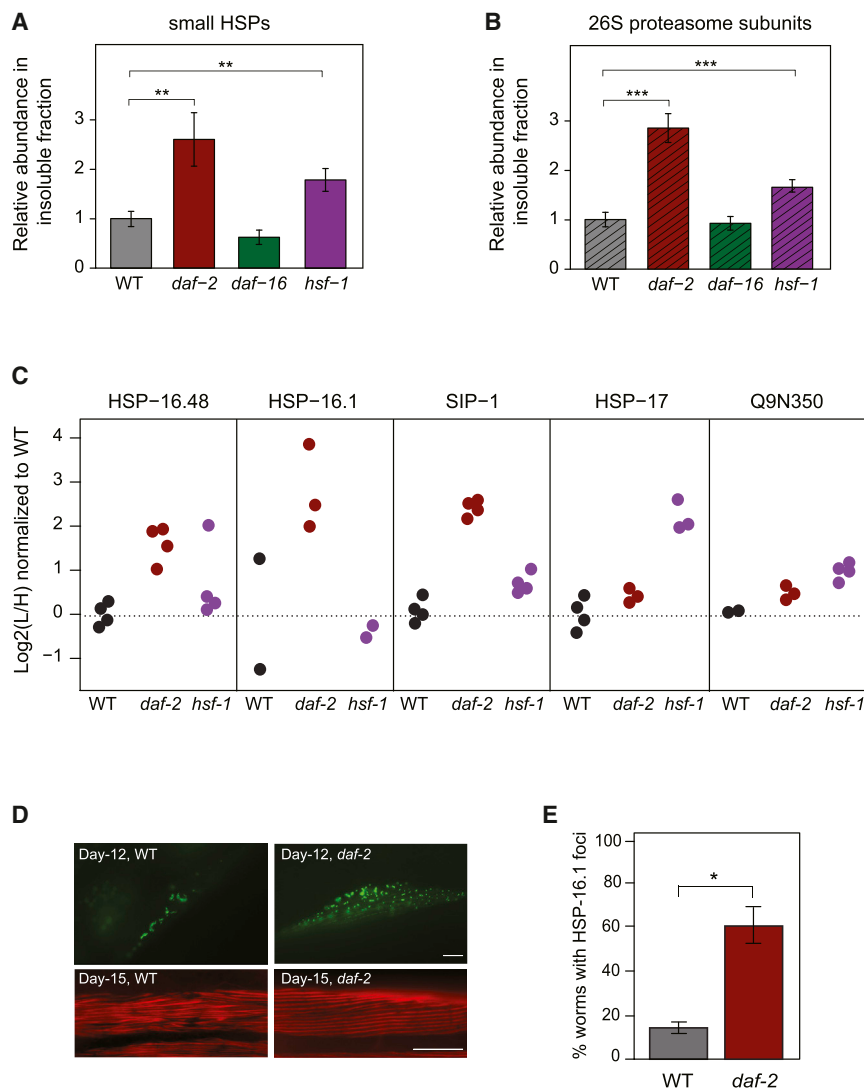


Figure 6. Aggregation of Small HSPs and Proteasome in Lifespan Mutant Worms

(A) Abundance of small HSPs in the insoluble fraction of *daf-2*, *daf-16*, and *hsf-1* mutant relative to WT animals as determined by summed absolute LFQ values. There were 6–11 different small HSPs that were quantified. **p value < 0.0075 (WT versus *daf-2*) and < 0.0022 (WT versus *hsf-1*) from Welch's t tests. Averages \pm SD are given for four replicate experiments.

(B) Abundance of 26S proteasome subunits in the insoluble fraction of *daf-2*, *daf-16*, and *hsf-1* mutant relative to WT animals. There were 19–27 subunits that were quantified. ***p < 2.1×10^{-4} (WT versus *daf-2*) and < 4.6×10^{-4} (WT versus *hsf-1*) from Welch's t tests. Averages \pm SD are given for four replicate experiments.

(C) Enrichment of the small HSPs HSP-16.1, HSP-16.48, SIP-1, HSP-17, and Q9N350 in the insoluble fractions of day 12 WT (black circles), *daf-2* mutant worms (red circles), and *hsf-1* mutant worms (purple circles). Data from two to four independent experiments are shown.

(D and E) Formation of HSP-16.1 inclusions in muscle cells is shown.

(D) Representative fluorescence images of muscle cells of WT and *daf-2* mutant animals expressing HSP-16.1::GFP (top). Actin was stained with rhodamine-phalloidin (bottom). Scale bar, 10 μ m. (E) Animals with HSP-16.1::GFP inclusions in muscle cells were quantified (20 animals per group). Averages \pm SD are given in % of total. *p < 0.01 from Welch's t test.

See also Figure S6 and Table S1.

aggregating in WT (Figures 5D–5G). These findings support the hypothesis of an extrinsic rescuing mechanism of aggregation that is activated in the *daf-2* mutant, modulating the intrinsic properties of proteins that typically govern aggregation. As a result, aggregation is enhanced for a set of proteins that have certain properties resembling disease-associated proteins with structural disorder (Knowles et al., 2014). By contrast, aggregation in the *hsf-1* mutant correlates with intrinsic aggregation scores (Figure 5D), consistent with a degeneration mechanism arising from the premature decline of proteostasis.

Among the proteins that were strongly increased in the insoluble fraction of *daf-2* mutant animals were several small HSPs (Figure 6A). These proteins contribute \sim 7% to total aggregate load, suggesting that they may be involved in a “protective aggregation response”. Small HSPs were also enriched in the insoluble fraction of *hsf-1* mutant animals, although to a lesser extent, but not in the aggregates of the *daf-16* mutant (Figure 6A). Besides small HSPs, 26S proteasome complexes were enriched in the insoluble fractions (Figure 6B), most

made the major contribution by mass to the aggregates in the *daf-2* mutant, while HSP-17 was most enriched in the aggregates of *hsf-1* mutant animals (Figure 6C; Table S1F). To monitor the behavior of HSP-16.1 during aging, we used strains expressing GFP-tagged HSP-16.1 (*hsp-16.1::gfp*) under its endogenous promoter. HSP-16.1-GFP formed inclusions in muscle cells. This phenomenon was strongly enhanced in *daf-2* mutant worms, with \sim 60% of animals at day 12 containing inclusions, compared to \sim 20% in WT (Figures 6D and 6E). While the actin architecture of muscle cells was well preserved in *daf-2* mutant animals, the muscles of day 15 WT animals showed a reduced structural integrity (Figure 6D). Indeed, the *daf-2* mutant animals displayed a higher proteostasis capacity, as reflected in their ability to maintain the metastable FlucDM-GFP (Gupta et al., 2011) expressed in muscle in a functionally active state. While similar levels of total and soluble FlucDM-GFP protein were present in day 12 WT and *daf-2* mutant worms, the latter contained \sim 4-fold more luciferase activity (Figure S6D). In contrast, a muscle specific poly-glutamine (polyQ) protein construct (Q35-GFP)

strongly in the *daf-2* mutant strain, but contributed only \sim 1% to total aggregate load.

Interestingly, the proportion of specific small HSPs in the aggregates differed between strains. SIP-1 and HSP-16.1

aggregated more extensively in *daf-2* mutant worms already early in adulthood (day 2), and semi-denaturing detergent agarose gel electrophoresis (SDD-AGE) of worm lysates revealed that the protein accumulated predominantly in higher molecular weight, SDS-resistant oligomers (Figure S6E).

Taken together, these results suggest that *daf-2* mutant animals drive a set of aberrant, potentially toxic proteins into insoluble aggregates, thereby sequestering them and improving overall proteostasis. Small HSPs are likely to play a role in this process.

DISCUSSION

Age-Dependent Deterioration of Proteome Balance

Organisms allocate considerable resources toward maintaining proteome composition, including the relative balance of subunits of multi-protein complexes (Li et al., 2014). Using quantitative mass spectrometric methods, we have shown here that aging in *C. elegans* is associated with the progressive failure to maintain protein homeostasis, resulting in extensive proteome remodeling and protein imbalances. These imbalances are largely due to changes at the level of protein translation and turnover and give rise to the accumulation of potentially harmful, aggregation-prone species (Figures 7A and 7B). Our analysis revealed that the sequestration of such proteins in insoluble aggregates is a protective strategy that contributes to maintaining proteome integrity during aging.

The extensive proteome remodeling during aging in *C. elegans* is contrary to observations in tissues of aged mice, where comparatively minor proteomic changes were detected with a similar experimental approach (Walther and Mann, 2011). Evidently, mammals devote greater resources to maintaining proteome balance, resulting in a more protracted proteostasis decline. These differences correlate with different reproductive strategies in worms and mice, in which the former display a more extensive, time-restrained burst of reproduction, followed by a rather rapid and massive decline of somatic functions. Future studies on a range of metazoans will be necessary to establish whether deterioration of proteome integrity during aging or proteome stability is more typical.

Changes in the Proteostasis System during Lifespan

We showed that aging in *C. elegans* affects multiple components of the proteostasis system, most prominently protein biosynthesis and protein degradation. A decrease in the levels of cytosolic and mitochondrial ribosomal proteins was accompanied by an overall reduction in protein synthesis. In contrast, we observed an increase in the abundance of proteasome subunits and a corresponding increase of *in vitro* proteasome activity. These changes may initiate as a response to the altered physiological requirements of the aging organism (Shore and Ruvkun, 2013), but ultimately may prove insufficient or even detrimental (Figure 7B). The reduction of the levels of cytosolic ribosomes was associated with a pronounced imbalance in the stoichiometry of ribosomal proteins. Thus, attenuating the translational machinery as an adaptive measure imparts the danger of dysregulation of the essential machines that ensure balance of the proteostasis network, which in turn may promote

aging. In contrast, the increase in proteasomal subunits is likely to represent an attempt of the organism to remove aberrant protein species. Whether this proteasome upregulation is effective *in vivo* is unclear, however, given that proteasome function is generally thought to decline as a result of aging and protein aggregation (Hipp et al., 2014).

Protein Aggregation and Lifespan Extension

Previous studies demonstrated the formation of insoluble protein aggregates in aged worms (David et al., 2010; Reis-Rodrigues et al., 2012). Here, we performed an in-depth quantitative analysis of aggregation along the lifespan of *C. elegans*. We found that aggregation is a proteome-wide process which initiates mainly after day 6 of adulthood. Highly abundant proteins are generally more soluble and display lower intrinsic aggregation-propensities than less abundant ones, as previously predicted (Tartaglia et al., 2009). However, this higher solubility is still not sufficient in the end, as abundant proteins make by far the major contribution by mass to the age-dependent aggregate load. Importantly, proteome remodeling acts as a driver of aggregation by raising the level of a subset of proteins beyond a critical solubility limit (supersaturation) (Ciryam et al., 2013) (Figure 7D).

While protein aggregation may be merely a consequence of declining proteostasis capacity, our results provide evidence that a protective aggregation response is also an important mechanism of the aging organism to improve proteostasis and mitigate the effects of proteome imbalance. We observed that long-lived *daf-2* mutant animals accumulate increasing amounts of insoluble proteins during aging and that such accumulation correlates with a more effective maintenance of proteome composition (Figure 7C). Whereas the proteins that aggregate most in the short-lived *hsf-1* mutant are predicted to be more aggregation-prone, the enhanced aggregation in the long-lived *daf-2* mutant is much less dependent on intrinsic properties: the proteins that are most enriched in the insoluble fraction have lower aggregation scores, are less hydrophobic, more charged, and contain more structural disorder, arguing for the existence of an active, proteome-wide mechanism in promoting aggregation. This conclusion is consistent with the view that soluble oligomers are the major proteotoxic species in neurodegenerative diseases and that their sequestration into insoluble aggregates reduces proteotoxicity (Arrasate et al., 2004; Cohen et al., 2006, 2009). Interestingly, several highly toxic disease-associated proteins are rich in disordered structure and have low overall hydrophobicity (Knowles et al., 2014; Vendruscolo et al., 2011), properties resembling those of the proteins with enhanced aggregation in the *daf-2* mutant. Indeed, a mechanism of aggregate deposition under regulation of the insulin signaling pathway has been proposed for disease-related protein species, such as toxic A β peptide (Cohen et al., 2006). However, that an overall protective aggregate response operates at the proteome-scale during aging was not anticipated.

We assume that this protective aggregation response is only partially activated during normal *C. elegans* aging. As a result, WT worms are expected to accumulate a larger soluble pool of aberrant, potentially toxic proteins than *daf-2* mutant animals, eventually exhausting the available chaperone capacity needed for protein folding and conformational maintenance and the

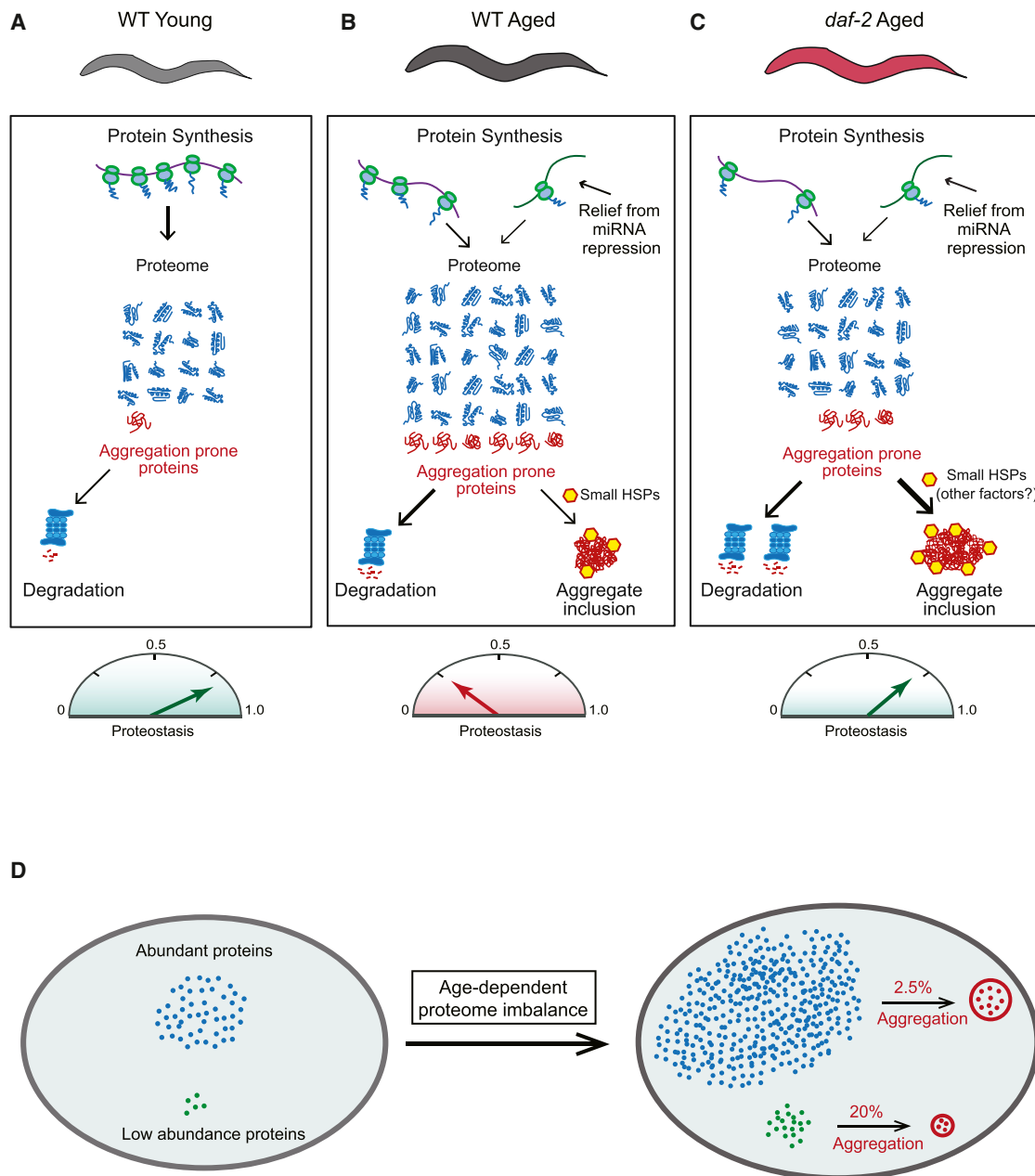


Figure 7. Proteome Maintenance during Aging in *C. elegans*

(A) The proteome of young adult WT worms is maintained in balance by the proteostasis system. Aberrant protein species, including metastable conformers and soluble aggregates (red) are efficiently cleared.

(B) In aged WT animals, numerous proteins increase in abundance and normal protein stoichiometries are lost due in part to a relief of miRNA-mediated translational repression. The amount of aggregation-prone species exceeds clearance capacity and insoluble aggregates associated with small HSPs accumulate. Mechanisms of protective aggregate formation are partially activated, and proteostasis is strongly reduced.

(C) Proteostasis collapse is delayed in aged *daf-2* mutant worms. Proteome imbalance and the soluble aggregate pool are reduced relative to age-matched WT animals, as clearance by protein degradation may be more effective and protective aggregate formation is fully activated.

(D) Protein aggregate loads increase proportionally to protein abundance. Although abundant proteins have lower aggregation propensities, they contribute more to aggregate load (see Figure 4). The age-dependent increase in expression level affects the subproteome of supersaturated proteins, which fail to maintain solubility as their levels increase and proteostasis capacity declines.

clearance of misfolded polypeptides (Figures 7B and 7C). Formation of insoluble aggregates may also be an, albeit insufficient, rescue attempt in the short-lived *hsf-1* mutant strain.

Among the proteostasis components with strongly enhanced, age-dependent insolubility were multiple small HSPs, a specific class of chaperones known to associate with aggregation-prone

proteins (Haslbeck et al., 2005). The small HSPs were most enriched in the insoluble fraction of *daf-2* mutant worms, consistent with the view that they may play a role as “extrinsic” promoters of aggregation. In support of this possibility, individual RNAi knockdown of several small HSPs, including SIP-1, caused a 25% shortening of lifespan in WT and *daf-2* mutant worms (Hsu et al., 2003) and overexpression resulted in lifespan extension (Walker and Lithgow, 2003). Having multi-valent binding ability for aberrant proteins, the small HSPs may act to seed and concentrate aggregate material, consistent with findings in vivo (Escusa-Toret et al., 2013; Kaganovich et al., 2008; Specht et al., 2011) and in vitro (Haslbeck et al., 2005; Jiao et al., 2005). The co-existence of multiple small HSPs suggests that different forms may vary in their structural specificity for endogenous proteins. Notably, small HSPs are also transcriptionally induced in the aging brain, while most other major chaperone components are downregulated (Brehme et al., 2014). The association of the 26S proteasome with the aggregates may also be functionally relevant. Although aggregates can inhibit the proteolytic activity of the proteasome (Andersson et al., 2013; Hipp et al., 2014), evidence has been presented that the ATPase chaperones of the 19S proteasome may promote aggregation (Rousseau et al., 2009).

Collectively, our data suggest that aging in *C. elegans* is associated with a progressive loss of proteome balance, which drives the accumulation of surplus and aberrant protein species that overburden the proteostasis system. As the maintenance of protein solubility imposes stringent constraints on proteome composition, effective aggregate management appears to be critical in determining lifespan.

EXPERIMENTAL PROCEDURES

C. elegans Strains and Growth Conditions

A list of strains used in this study is provided in the Extended Experimental Procedures. The Bristol strain N2 was used as WT. The L4 larval stage was considered as day 0. Bacterial cultures (ET505) for SILAC labeling were grown in $^{13}\text{C}_6$ - $^{15}\text{N}_2$ -lysine (heavy lysine) containing M63 minimal media (Krijgsveld et al., 2003) (see Extended Experimental Procedures for details).

Sample Preparation for Total Proteome Measurements

Briefly, worms were suspended in lysis buffer (4% SDS, 0.1 M Tris/HCl pH 8.0, and 1 mM EDTA), incubated at 95°C for 5 min, and further treated by ultrasonication. Typically, an aliquot of lysate containing 40 µg of protein was mixed with an equal amount of a heavy lysine labeled lysate pool (Figure 1A). Proteins were reduced, alkylated, and digested with endoproteinase LysC using the filter-aided sample preparation (FASP) method (Wiśniewski et al., 2009). Peptide mixtures were either analyzed without fractionation or after fractionation by isoelectric focusing, as described in the Extended Experimental Procedures.

Isolation of Protein Aggregates

Worms were resuspended in lysis buffer (50 mM Tris/HCl pH 8.0, 0.5 M NaCl, 4 mM EDTA, 1% volume/volume (v/v) Igepal CA630, and complete protease inhibitor cocktail; Roche Diagnostics), disrupted by sonication, and clarified by low-speed centrifugation (1 min, 1,000 relative centrifugal force [rcf]). Insoluble proteins were sedimented by ultracentrifugation (500,000 rcf at 4°C, 10 min), washed twice with lysis buffer containing 0.15 M NaCl and 0.5% sodium deoxycholate, and solubilized in SDS sample buffer for 10 min at 95°C. For quantitative proteome measurements of aggregated proteins, an aliquot of pooled total lysate from heavy lysine labeled animals was added prior to ultracentrifugation. For experiments measuring aggregation propen-

sities, unlabeled worm lysates were first fractionated and subsequently supplemented with SILAC-labeled whole cell lysate.

MS and Data Analysis

Peptides were separated by reversed phase nano-high-performance liquid chromatography (HPLC) and sprayed online into LTQ-Orbitrap Velos or Orbitrap-Elite mass spectrometers (Thermo Fisher). In each scan cycle, fragmentation spectra of the ten most intense peptide precursors in the survey scan were acquired in the higher-energy collisional dissociation (HCD) mode. Raw data were processed using the MaxQuant software environment (Cox and Mann, 2008) and peak lists were searched with Andromeda (Cox et al., 2011) against a database containing the translation of all predicted proteins listed in UniProt (release January 15, 2012), as well as a list of commonly observed contaminants and the National Center for Biotechnology Information (NCBI) protein database of *E. coli* strain K12. The minimal required peptide length was set to seven amino acids and both protein and peptide identifications were accepted at a false discovery rate of 1%. To identify aggregation-prone proteins that were significantly affected by aging, those proteins that were quantified in at least three out of four biological replicate experiments at day 1 and day 12 were subjected to a Welch's t test and filtered based on a 5% permutation-based false discovery rate threshold.

Miscellaneous

Proteasome activity assays, detection of polyQ aggregates by SDD-AGE, light microscopy, and methods used for bioinformatic analyses are described in the Extended Experimental Procedures.

ACCESSION NUMBERS

Proteomics raw data and selected MaxQuant output files have been deposited to the ProteomeXchange Consortium via the PRIDE partner repository with the data set identifier PXD001364.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Extended Experimental Procedures, six figures, and four tables and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.03.032>.

AUTHOR CONTRIBUTIONS

F.U.H., M.M., D.M.W., and P.K. conceived the approach. D.M.W. and P.K. designed and performed the experiments with contributions from M.Z. M.M. supervised the proteomics analyses. D.M.W. and S.P. performed bioinformatics analyses with contributions from P.K. G.V. and P.C. analyzed the protein aggregation data and interpreted the results together with M.V., C.M.D., and R.I.M. F.U.H., D.M.W., and P.K. wrote the manuscript with contributions from M.M., M.V., C.M.D., and R.I.M.

ACKNOWLEDGMENTS

We thank the Caenorhabditis Genetics Center for providing most of the strains used in this study. *hsp-16.1::gfp* transgenic worms were provided by Dr. Junho Lee, Seoul National University. This work was supported by the European Commission under FP7 GA number ERC-2012-SyG_318987-Toxic Protein Aggregation in Neurodegeneration (ToPAG), the Wellcome Trust (094425/Z/10/Z), the Center for Integrated Protein Science Munich (CIPSM), and the Munich Cluster for Systems Neurology (SyNergy). We would further like to thank Igor Paron and Korbinian Mayr for excellent technical assistance with MS instrumentation, as well as Daniel Hornburg, Manajit Hayer-Hartl, and members of the Hartl laboratory for critically reading this manuscript.

Received: October 23, 2014

Revised: January 12, 2015

Accepted: February 24, 2015

Published: May 7, 2015

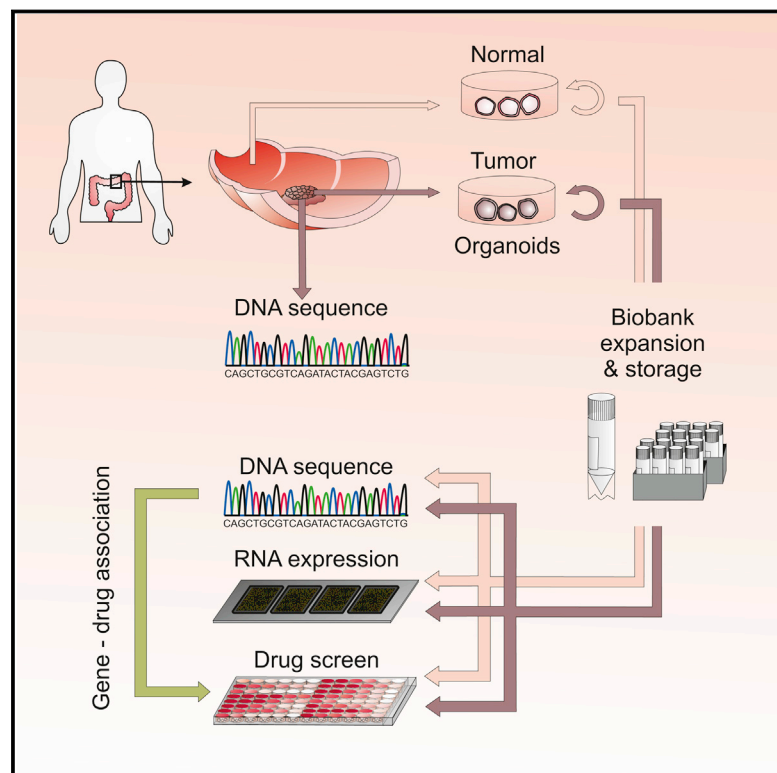
REFERENCES

- Andersson, V., Hanzén, S., Liu, B., Molin, M., and Nyström, T. (2013). Enhancing protein disaggregation restores proteasome activity in aged cells. *Aging (Albany, N.Y. Online)* 5, 802–812.
- Arrasate, M., Mitra, S., Schweitzer, E.S., Segal, M.R., and Finkbeiner, S. (2004). Inclusion body formation reduces levels of mutant huntingtin and the risk of neuronal death. *Nature* 431, 805–810.
- Balch, W.E., Morimoto, R.I., Dillin, A., and Kelly, J.W. (2008). Adapting proteostasis for disease intervention. *Science* 319, 916–919.
- Ben-Zvi, A., Miller, E.A., and Morimoto, R.I. (2009). Collapse of proteostasis represents an early molecular event in *Caenorhabditis elegans* aging. *Proc. Natl. Acad. Sci. USA* 106, 14914–14919.
- Bensimon, A., Heck, A.J., and Aebersold, R. (2012). Mass spectrometry-based proteomics and network biology. *Annu. Rev. Biochem.* 81, 379–405.
- Brehme, M., Voisine, C., Rolland, T., Wachi, S., Soper, J.H., Zhu, Y., Orton, K., Villella, A., Garza, D., Vidal, M., et al. (2014). A chaperome subnetwork safeguards proteostasis in aging and neurodegenerative disease. *Cell Rep.* 9, 1135–1150.
- Budovskaya, Y.V., Wu, K., Southworth, L.K., Jiang, M., Tedesco, P., Johnson, T.E., and Kim, S.K. (2008). An elt-3/elt-5/elt-6 GATA transcription circuit guides aging in *C. elegans*. *Cell* 134, 291–303.
- Chikina, M.D., Huttenhower, C., Murphy, C.T., and Troyanskaya, O.G. (2009). Global prediction of tissue-specific gene expression and context-dependent gene networks in *Caenorhabditis elegans*. *PLoS Comput. Biol.* 5, e1000417.
- Ciryam, P., Tartaglia, G.G., Morimoto, R.I., Dobson, C.M., and Vendruscolo, M. (2013). Widespread aggregation and neurodegenerative diseases are associated with supersaturated proteins. *Cell Rep.* 5, 781–790.
- Cohen, E., Bieschke, J., Perciavalle, R.M., Kelly, J.W., and Dillin, A. (2006). Opposing activities protect against age-onset proteotoxicity. *Science* 313, 1604–1610.
- Cohen, E., Paulsson, J.F., Blinder, P., Burstyn-Cohen, T., Du, D., Estepa, G., Adame, A., Pham, H.M., Holzenberger, M., Kelly, J.W., et al. (2009). Reduced IGF-1 signaling delays age-associated proteotoxicity in mice. *Cell* 139, 1157–1169.
- Cox, J., and Mann, M. (2008). MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat. Biotechnol.* 26, 1367–1372.
- Cox, J., and Mann, M. (2011). Quantitative, high-resolution proteomics for data-driven systems biology. *Annu. Rev. Biochem.* 80, 273–299.
- Cox, J., Neuhauser, N., Michalski, A., Scheltema, R.A., Olsen, J.V., and Mann, M. (2011). Andromeda: a peptide search engine integrated into the MaxQuant environment. *J. Proteome Res.* 10, 1794–1805.
- David, D.C., Ollikainen, N., Trinidad, J.C., Cary, M.P., Burlingame, A.L., and Kenyon, C. (2010). Widespread protein aggregation as an inherent part of aging in *C. elegans*. *PLoS Biol.* 8, e1000450.
- Demontis, F., and Perrimon, N. (2010). FOXO/4E-BP signaling in *Drosophila* muscles regulates organism-wide proteostasis during aging. *Cell* 143, 813–825.
- Dong, M.Q., Venable, J.D., Au, N., Xu, T., Park, S.K., Cociorva, D., Johnson, J.R., Dillin, A., and Yates, J.R., 3rd. (2007). Quantitative mass spectrometry identifies insulin signaling targets in *C. elegans*. *Science* 317, 660–663.
- Douglas, P.M., and Dillin, A. (2010). Protein homeostasis and aging in neurodegeneration. *J. Cell Biol.* 190, 719–729.
- Escusa-Toret, S., Vonk, W.I., and Frydman, J. (2013). Spatial sequestration of misfolded proteins by a dynamic chaperone pathway enhances cellular fitness during stress. *Nat. Cell Biol.* 15, 1231–1243.
- Finkel, T., and Holbrook, N.J. (2000). Oxidants, oxidative stress and the biology of ageing. *Nature* 408, 239–247.
- Gidalevitz, T., Ben-Zvi, A., Ho, K.H., Brignull, H.R., and Morimoto, R.I. (2006). Progressive disruption of cellular protein folding in models of polyglutamine diseases. *Science* 311, 1471–1474.
- Golden, T.R., and Melov, S. (2004). Microarray analysis of gene expression with age in individual nematodes. *Aging Cell* 3, 111–124.
- Gupta, R., Kasturi, P., Bracher, A., Loew, C., Zheng, M., Villella, A., Garza, D., Hartl, F.U., and Raychaudhuri, S. (2011). Firefly luciferase mutants as sensors of proteome stress. *Nat. Methods* 8, 879–884.
- Hartl, F.U., Bracher, A., and Hayer-Hartl, M. (2011). Molecular chaperones in protein folding and proteostasis. *Nature* 475, 324–332.
- Haslbeck, M., Franzmann, T., Weinfurtner, D., and Buchner, J. (2005). Some like it hot: the structure and function of small heat-shock proteins. *Nat. Struct. Mol. Biol.* 12, 842–846.
- Hipp, M.S., Park, S.H., and Hartl, F.U. (2014). Proteostasis impairment in protein-misfolding and -aggregation diseases. *Trends Cell Biol.* 24, 506–514.
- Hsu, A.-L., Murphy, C.T., and Kenyon, C. (2003). Regulation of aging and age-related disease by DAF-16 and heat-shock factor. *Science* 300, 1142–1145.
- Ibáñez-Ventoso, C., Yang, M., Guo, S., Robins, H., Padgett, R.W., and Driscoll, M. (2006). Modulated microRNA expression during adult lifespan in *Caenorhabditis elegans*. *Aging Cell* 5, 235–246.
- Jiao, W., Li, P., Zhang, J., Zhang, H., and Chang, Z. (2005). Small heat-shock proteins function in the insoluble protein complex. *Biochem. Biophys. Res. Commun.* 335, 227–231.
- Kaganovich, D., Kopito, R., and Frydman, J. (2008). Misfolded proteins partition between two distinct quality control compartments. *Nature* 454, 1088–1095.
- Kenyon, C., Chang, J., Gensch, E., Rudner, A., and Tabtiang, R. (1993). A *C. elegans* mutant that lives twice as long as wild type. *Nature* 366, 461–464.
- Kirstein-Miles, J., Scior, A., Deuerling, E., and Morimoto, R.I. (2013). The nascent polypeptide-associated complex is a key regulator of proteostasis. *EMBO J.* 32, 1451–1468.
- Knowles, T.P., Vendruscolo, M., and Dobson, C.M. (2014). The amyloid state and its association with protein misfolding diseases. *Nat. Rev. Mol. Cell Biol.* 15, 384–396.
- Krijgsvelde, J., Ketting, R.F., Mahmoudi, T., Johansen, J., Artal-Sanz, M., Verrijzer, C.P., Plasterk, R.H.A., and Heck, A.J.R. (2003). Metabolic labeling of *C. elegans* and *D. melanogaster* for quantitative proteomics. *Nat. Biotechnol.* 21, 927–931.
- Kumar, L., and Futschik, M. (2007). Mfuzz: a software package for soft clustering of microarray data. *Bioinformatics* 2, 5–7.
- Larance, M., Bailly, A.P., Pourkarimi, E., Hay, R.T., Buchanan, G., Coulthurst, S., Xirodimas, D.P., Gartner, A., and Lamond, A.I. (2011). Stable-isotope labeling with amino acids in nematodes. *Nat. Methods* 8, 849–851.
- Li, G.W., Burkhardt, D., Gross, C., and Weissman, J.S. (2014). Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. *Cell* 157, 624–635.
- McElwee, J., Bubb, K., and Thomas, J.H. (2003). Transcriptional outputs of the *Caenorhabditis elegans* forkhead protein DAF-16. *Aging Cell* 2, 111–121.
- Morley, J.F., and Morimoto, R.I. (2004). Regulation of longevity in *Caenorhabditis elegans* by heat shock factor and molecular chaperones. *Mol. Biol. Cell* 15, 657–664.
- Morley, J.F., Brignull, H.R., Weyers, J.J., and Morimoto, R.I. (2002). The threshold for polyglutamine-expansion protein aggregation and cellular toxicity is dynamic and influenced by aging in *Caenorhabditis elegans*. *Proc. Natl. Acad. Sci. USA* 99, 10417–10422.
- Murphy, C.T., McCarroll, S.A., Bargmann, C.I., Fraser, A., Kamath, R.S., Ahringer, J., Li, H., and Kenyon, C. (2003). Genes that act downstream of DAF-16 to influence the lifespan of *Caenorhabditis elegans*. *Nature* 424, 277–283.
- Olzscha, H., Schermann, S.M., Woerner, A.C., Pinkert, S., Hecht, M.H., Tartaglia, G.G., Vendruscolo, M., Hayer-Hartl, M., Hartl, F.U., and Vabulas, R.M. (2011). Amyloid-like aggregates sequester numerous metastable proteins with essential cellular functions. *Cell* 144, 67–78.
- Ong, S.E., Blagoev, B., Kratchmarova, I., Kristensen, D.B., Steen, H., Pandey, A., and Mann, M. (2002). Stable isotope labeling by amino acids in cell culture,

- SILAC, as a simple and accurate approach to expression proteomics. *Mol. Cell. Proteomics* 1, 376–386.
- Oromendia, A.B., Dodgson, S.E., and Amon, A. (2012). Aneuploidy causes proteotoxic stress in yeast. *Genes Dev.* 26, 2696–2708.
- Prahlad, V., and Morimoto, R.I. (2009). Integrating the stress response: lessons for neurodegenerative diseases from *C. elegans*. *Trends Cell Biol.* 19, 52–61.
- Reis-Rodrigues, P., Czerwieniec, G., Peters, T.W., Evani, U.S., Alavez, S., Gaman, E.A., Vantipalli, M., Mooney, S.D., Gibson, B.W., Lithgow, G.J., and Hughes, R.E. (2012). Proteomic analysis of age-dependent changes in protein solubility identifies genes that modulate lifespan. *Aging Cell* 11, 120–127.
- Rousseau, E., Kojima, R., Hoffner, G., Djian, P., and Bertolotti, A. (2009). Misfolding of proteins with a polyglutamine expansion is facilitated by proteasomal chaperones. *J. Biol. Chem.* 284, 1917–1929.
- Schwanhäusser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W., and Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature* 473, 337–342.
- Shore, D.E., and Ruvkun, G. (2013). A cytoprotective perspective on longevity regulation. *Trends Cell Biol.* 23, 409–420.
- Sormanni, P., Aprile, F.A., and Vendruscolo, M. (2015a). The CamSol method of rational design of protein mutants with enhanced solubility. *J. Mol. Biol.* 427, 478–490.
- Sormanni, P., Camilloni, C., Fariselli, P., and Vendruscolo, M. (2015b). The s2D method: simultaneous sequence-based prediction of the statistical populations of ordered and disordered regions in proteins. *J. Mol. Biol.* 427, 982–996.
- Specht, S., Miller, S.B., Mogk, A., and Bukau, B. (2011). Hsp42 is required for sequestration of protein aggregates into deposition sites in *Saccharomyces cerevisiae*. *J. Cell Biol.* 195, 617–629.
- Stingle, S., Stoehr, G., Peplowska, K., Cox, J., Mann, M., and Storchova, Z. (2012). Global analysis of genome, transcriptome and proteome reveals the response to aneuploidy in human cells. *Mol. Syst. Biol.* 8, 608.
- Tartaglia, G.G., Pawar, A.P., Campioni, S., Dobson, C.M., Chiti, F., and Vendruscolo, M. (2008). Prediction of aggregation-prone regions in structured proteins. *J. Mol. Biol.* 380, 425–436.
- Tartaglia, G.G., Pechmann, S., Dobson, C.M., and Vendruscolo, M. (2009). A relationship between mRNA expression levels and protein solubility in *E. coli*. *J. Mol. Biol.* 388, 381–389.
- Taylor, R.C., and Dillin, A. (2013). XBP-1 is a cell-nonautonomous regulator of stress resistance and longevity. *Cell* 153, 1435–1447.
- van Oosten-Hawle, P., and Morimoto, R.I. (2014). Organismal proteostasis: role of cell-nonautonomous regulation and transcellular chaperone signaling. *Genes Dev.* 28, 1533–1543.
- Vendruscolo, M., Knowles, T.P., and Dobson, C.M. (2011). Protein solubility and protein homeostasis: a generic view of protein misfolding disorders. *Cold Spring Harb. Perspect. Biol.* 3, a010454.
- Vilchez, D., Morante, I., Liu, Z., Douglas, P.M., Merkwirth, C., Rodrigues, A.P., Manning, G., and Dillin, A. (2012). RPN-6 determines *C. elegans* longevity under proteotoxic stress conditions. *Nature* 489, 263–268.
- Walker, G.A., and Lithgow, G.J. (2003). Lifespan extension in *C. elegans* by a molecular chaperone dependent upon insulin-like signals. *Aging Cell* 2, 131–139.
- Walther, D.M., and Mann, M. (2011). Accurate quantification of more than 4000 mouse tissue proteins reveals minimal proteome changes during aging. *Mol. Cell. Proteomics* 10, M110.004523.
- Welker, N.C., Habig, J.W., and Bass, B.L. (2007). Genes misregulated in *C. elegans* deficient in Dicer, RDE-4, or RDE-1 are enriched for innate immunity genes. *RNA* 13, 1090–1102.
- Wiśniewski, J.R., Zougman, A., Nagaraj, N., and Mann, M. (2009). Universal sample preparation method for proteome analysis. *Nat. Methods* 6, 359–362.

Prospective Derivation of a Living Organoid Biobank of Colorectal Cancer Patients

Graphical Abstract



Authors

Marc van de Wetering,
Hayley E. Francies, ..., Mathew J. Garnett,
Hans Clevers

Correspondence

mg12@sanger.ac.uk (M.J.G.),
h.clevers@hubrecht.eu (H.C.)

In Brief

3D organoid cultures derived from healthy and tumor tissue from colorectal cancer patients are used for a high throughput drug screen to identify gene-drug associations that may facilitate personalized therapy.

Highlights

- Tumor and normal organoids were derived from colorectal carcinoma patients
- Tumor organoids recapitulate somatic copy number and mutation spectra found in CRC
- Organoids are amenable to high-throughput drug screening
- Patient-derived organoids allow personalized therapy design

Accession Numbers

GSE64392
GSE65253



Prospective Derivation of a Living Organoid Biobank of Colorectal Cancer Patients

Marc van de Wetering,^{1,2,12} Hayley E. Francies,^{3,12} Joshua M. Francis,^{4,5,12} Gergana Bounova,⁶ Francesco Iorio,⁷ Apollo Pronk,⁸ Winan van Houdt,⁸ Joost van Gorp,⁹ Amaro Taylor-Weiner,⁴ Lennart Kester,¹ Anne McLaren-Douglas,³ Joyce Blokker,^{1,2} Sridevi Jaksani,^{1,2} Sina Bartfeld,¹ Richard Volckman,¹⁰ Peter van Sluis,¹⁰ Vivian S.W. Li,¹¹ Sara Seepo,⁴ Chandra Sekhar Pedamallu,^{4,5} Kristian Cibulskis,⁴ Scott L. Carter,^{4,5} Aaron McKenna,⁴ Michael S. Lawrence,⁴ Lee Lichtenstein,⁴ Chip Stewart,⁴ Jan Koster,¹⁰ Rogier Versteeg,¹⁰ Alexander van Oudenaarden,¹ Julio Saez-Rodriguez,⁷ Robert G.J. Vries,^{1,2} Gad Getz,^{4,5} Lodewyk Wessels,⁶ Michael R. Stratton,³ Ultan McDermott,³ Matthew Meyerson,^{4,5} Mathew J. Garnett,^{3,*} and Hans Clevers^{1,2,*}

¹Hubrecht Institute, Royal Netherlands Academy of Arts and Sciences (KNAW), Cancer Genomics and University Medical Center, 3584 CT Utrecht, the Netherlands

²Foundation Hubrecht Organoid Technology (HUB), 3584 CT Utrecht, the Netherlands

³Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire CB10 1SA, UK

⁴The Broad Institute of MIT and Harvard, Cambridge, MA 02142, USA

⁵Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MA 02115, USA

⁶Computational Cancer Biology, Netherlands Cancer Institute, 1066 CX Amsterdam, the Netherlands

⁷European Molecular Biology Laboratory, European Bioinformatics Institute, Wellcome Trust Genome Campus, Cambridge CB10 1SA, UK

⁸Department of Surgery, Diaconessenhuis, 3582 KE Utrecht, the Netherlands

⁹Department of Pathology, Diaconessenhuis, 3582 KE Utrecht, the Netherlands

¹⁰Department of Oncogenomics, Academic Medical Center, University of Amsterdam, 1105 AZ Amsterdam, the Netherlands

¹¹Division of Stem Cell Biology and Developmental Genetics, MRC National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK

¹²Co-first author

*Correspondence: mg12@sanger.ac.uk (M.J.G.), h.clevers@hubrecht.eu (H.C.)

<http://dx.doi.org/10.1016/j.cell.2015.03.053>

SUMMARY

In R-spondin-based 3D cultures, Lgr5 stem cells from multiple organs form ever-expanding epithelial organoids that retain their tissue identity. We report the establishment of tumor organoid cultures from 20 consecutive colorectal carcinoma (CRC) patients. For most, organoids were also generated from adjacent normal tissue. Organoids closely recapitulate several properties of the original tumor. The spectrum of genetic changes within the “living biobank” agrees well with previous large-scale mutational analyses of CRC. Gene expression analysis indicates that the major CRC molecular subtypes are represented. Tumor organoids are amenable to high-throughput drug screens allowing detection of gene-drug associations. As an example, a single organoid culture was exquisitely sensitive to Wnt secretion (porcupine) inhibitors and carried a mutation in the negative Wnt feedback regulator *RNF43*, rather than in *APC*. Organoid technology may fill the gap between cancer genetics and patient trials, complement cell-line- and xenograft-based drug studies, and allow personalized therapy design.

INTRODUCTION

Colorectal carcinoma (CRC) represents one of the major forms of cancer. Seminal studies have revealed a series of molecular

pathways that are critical to the pathogenesis of CRC, including WNT, RAS-MAPK, PI3K, P53, TGF- β , and DNA mismatch repair (Fearon, 2011; Fearon and Vogelstein, 1990). Large-scale sequencing analyses have dramatically extended the list of recurrently mutated genes and chromosomal translocations (Garraway and Lander, 2013; Vogelstein et al., 2013). CRC cases are characterized by either microsatellite instability (MSI) (associated with a hyper-mutator phenotype), or as microsatellite-stable (MSS) but chromosomally unstable (CIN) (Lengauer et al., 1997). The absolute number and combination of genetic alterations in CRC confounds our ability to unravel the functional contribution of each of these potential cancer genes. Thus, while genome changes in tumors of individual patients can be assessed in great detail and at low cost, these data are difficult to interpret in terms of prognosis, drug response, or patient outcome, necessitating model systems for analysis of genotype-to-phenotype correlations.

Self-renewal of the intestinal epithelium is driven by Lgr5 stem cells located in crypts (Barker et al., 2007). We have recently developed a long-term culture system that maintains basic crypt physiology (Sato et al., 2009). Wnt signals are required for the maintenance of active crypt stem cells (Korinek et al., 1998; Kuhnert et al., 2004; Pinto et al., 2003). Indeed, the Wnt agonist R-spondin1 induces dramatic crypt hyperplasia in vivo (Kim et al., 2005). R-spondin-1 is the ligand for Lgr5 (Carmon et al., 2011; de Lau et al., 2011). Epidermal growth factor (EGF) signaling is associated with intestinal proliferation (Wong et al., 2012), while transgenic expression of Noggin induces a dramatic increase in crypt numbers (Haramis et al., 2004). The combination of R-spondin-1, EGF, and Noggin in

Basement Membrane Extract (BME) sustains ever-expanding small intestinal organoids, which display all hallmarks of the original tissue in terms of architecture, cell-type composition, and self-renewal dynamics. We adapted the culture condition for long-term expansion of human colonic epithelium and primary colonic adenocarcinoma, by adding nicotinamide, A83-01 (Alk inhibitor), Prostaglandin E2, and the p38 inhibitor SB202190 (Sato et al., 2011). Of note, a 2D culture method for cells from normal and malignant primary tissue has been described by Liu et al. (2012).

Here, we explore organoid technology to routinely establish and phenotypically annotate “paired organoids” derived from adjacent tumor and healthy epithelium from CRC patients.

RESULTS

Establishment of a Living CRC Biobank

Surgically resected tissue was obtained from previously untreated CRC patients. Tissue from rectal cancer patients was excluded because they routinely undergo irradiation before surgery. For multiple tissues, we observe that normal tissue-derived organoids outcompete tumor organoids under the optimized culture conditions, presumably due to genomic instability and resulting apoptosis in the latter. Combination of Wnt3A and the Wnt amplifier R-spondin1 is essential to grow organoids from normal epithelium. Over 90% of CRC cases harbor mutations that aberrantly activate the Wnt signaling pathway (Cancer Genome Atlas Network, 2012), so we exploited the Wnt-dependency of normal colonic stem cells to selectively expand tumor organoids. A total of 22 tumor organoid cultures and 19 normal-adjacent organoid cultures were derived from 20 patients (P19 and P24 each carried two primary tumors separated by >10 cm; Figure 1A). We successfully generated organoid cultures from 22 of 27 tumor samples. For one, we never observed growth. Four were lost due to bacterial/yeast infection. Since then, we have added next-generation antibiotics (see Experimental Procedures) and currently observe an ~90% success rate.

The number of primary tumor organoids varied between patient samples, with some tumors rendering thousands of primary organoids whereas others yielded only 10–20 primary organoids. This difference in derivation likely reflects the heterogeneous composition of tumors, with proliferative areas intermingled with regions of differentiated cells, stromal cells or necrosis. The growth rate of the organoids from patients 5 and 27 decreased over time, which prohibited their inclusion in the drug screen. All other organoids could be readily expanded and frozen to create a master cell bank. Upon thawing, cell survival was typically >80%. Unlike healthy tissue-derived organoids, tumor-derived organoids presented with a range of patient-specific morphologies, ranging from thin-walled cystic structures to compact organoids devoid of a lumen. H&E staining on primary tumors and the corresponding organoids revealed that the “cystic versus solid”-organization of the epithelium was generally preserved. Yet, marker expression analysis (KI67, OLFM4, KRT 20, Alcian blue) revealed heterogeneity both between patients and individual organoids within each culture (Figure 1B; Data S1).

Genomic Characterization of Tumor-Derived Organoids

Genomic DNA was isolated from tumor and matched normal organoid cultures for whole-exome sequencing in order to identify tumor-specific somatic mutations (Cancer Genome Atlas Network, 2012). Genomic DNA from the corresponding biopsy specimens were available for comparative analysis for 16 of these cases (Table S1A). The mutation rates per Mb varied widely for different tumor organoids (range 2.0–77.9), with a median value of 3.7 in the tumor organoids, similar to the median rate of 3.6 in the biopsy samples (Figure 2A; Table S1B). Mutations were predominantly CpG to T transitions, consistent with results from large-scale CRC sequencing (Figures S1A and S1B; Table S1C). Of the 22 tumor organoids, six displayed hypermutation (>10 mutations/Mb): P7, P10 and the organoids from the two patients with two tumors each (P19a and P19b, P24a and P24b). Interestingly, the P19a and P19b tumors share TP53 R273C and BRAF V600E alterations, suggesting they arose from the same somatically altered progenitor cell but then diverged to acquire independent secondary alterations (Figures S1C and S1D). In contrast, the P24a and P24b tumors share 80% (469/590) of somatic alterations but then have discordant driving alterations in *APC* and *TP53*, indicating that the hypermutator phenotype may have been present prior to the acquisition of growth promoting mutations (Figures S1E and S1F). The frequency of hypermutated organoid cultures in our patient panel (20%; 4 of 20) agreed with the reported frequency in a much larger cohort of clinical samples and display comparable somatic copy number alterations (SCNAs) (Figure 2B; Table S1D) (Bass et al., 2011; Cancer Genome Atlas Network, 2012). The successful derivation of both hypermutated and non-hypermutated organoids implies an absence of culture-based bias.

Somatic variants within the coding regions in organoid cultures were highly concordant with the corresponding biopsy specimen for both hypermutated and non-hypermutated patients (median = 0.88 frequency of concordance, range 0.62–1.00) (Figure 3A; Table S1E). Indeed, combined analysis of SCNAs and single nucleotide variants (SNVs) to infer Cancer Cell Fractions (CCF) (Carter et al., 2012; Landau et al., 2013) in the biopsy and tumor organoids, revealed that the common CRC driver mutations were maintained in culture. In 13 out of 14 organoid-biopsy pairs tested, tumor subclones sharing common CRC drivers were detected in the biopsy. In 50% of the organoids, a dominant subclone from the biopsy was present, likely representing sampling during derivation but it could also indicate loss in culture (Figures S2A and S2B; Tables S1F and S1G). Transcriptome analysis of single organoids showed subtle differences in gene expression within an organoid culture, confirming their heterogeneous composition. The differences in overall gene expression were more pronounced in the organoids derived from the hypermutant tumors (Figure S2C).

Discordant mutations were assessed for their likely biological significance in cancer, based on Cancer Gene Census and data reported from the PanCancer analysis of 5,000 whole exomes (Futreal et al., 2004; Lawrence et al., 2014). Only 4% (27/679) of discordant mutations found in organoids affected cancer-related genes, including a third hit to *APC*, which was already biallelically inactivated in P14, *SMAD4* mutation in P16, and *POLE* mutation in P19b (Table S1H). Cancer-significant genes

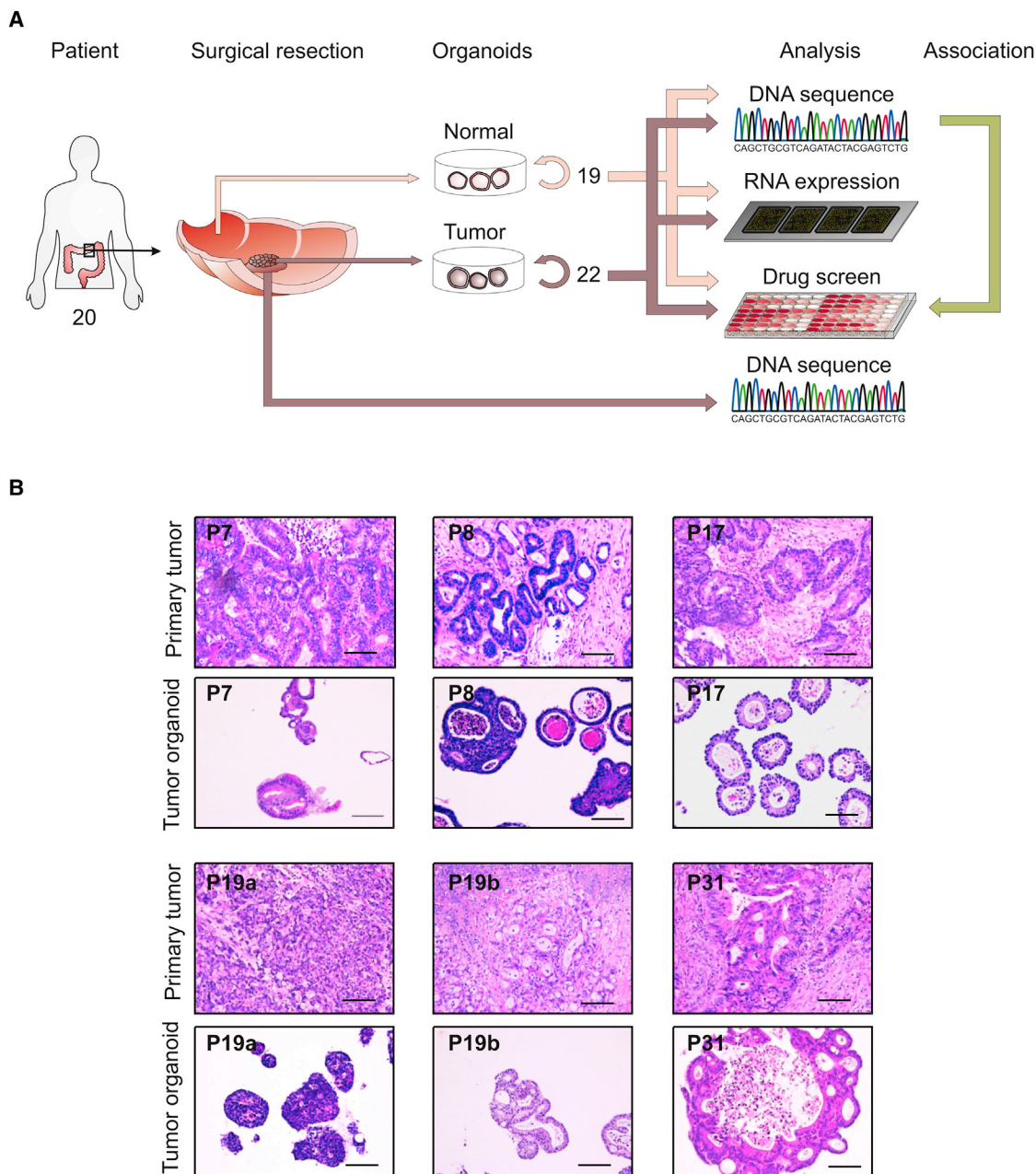


Figure 1. Derivation of Organoids from Primary Tissue

(A) Overview of the procedure. A total of 22 tumor organoids and 19 normal control organoids were derived and analyzed by exome-sequencing, RNA expression analysis and high-throughput drug screening. To determine the concordance between tumor organoids and primary tumor, DNA from the primary tumor was also isolated.

(B) Organoids architecture resembles primary tumor epithelium. H&E staining of primary tumor and the tumor organoids derived of these. A feature of most organoids is the presence of one or more lumens, resembling the tubular structures of the primary tumor (e.g., P8 and P19b). Tumors devoid of lumen give rise to compact organoids without lumen (P19a). Scale bar, 100 μ M.

See also [Data S1](#).

that were discordant in the biopsy represented 4.4% (12/271) ([Table S1H](#)). The discordant mutations had a mean allelic frequency of 10.3% and 34.1% for the biopsy and organoids, respectively. This could represent the enrichment or depletion of a sub-clonal population in the organoid culture present within

the original tumor, as well as acquisition of additional mutations during derivation or propagation.

The most commonly altered genes in CRC ([Bass et al., 2011](#); [Cancer Genome Atlas Network, 2012](#); [Lawrence et al., 2014](#)) were well represented in the organoid cultures ([Figure 3B](#);

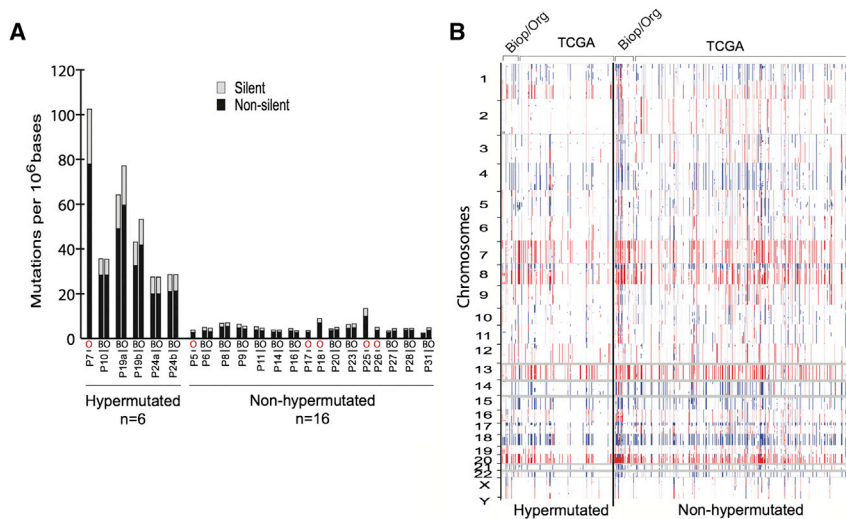


Figure 2. CRC Subtypes Are Present in Organoid Cultures

(A) Whole exome sequencing of the tumor and corresponding biopsy, when available, revealed the presence of hypermutated (>10 mutations/Mb) and non-hypermutated subtypes within the organoids. Comparable rates of mutations were observed in the tumor organoid (O) and tumor biopsy (B). Organoids without corresponding biopsy are indicated in with red (O).

(B) Comparison of somatic copy-number alterations found in the biopsies and corresponding organoids (Biop/Org) and TCGA CRC in hypermutated and non-hypermutated samples.

See also Figure S1 and Tables S1A–S1D.

Tables S1I and S1J). Inactivating alterations to the tumor suppressors *APC*, *TP53*, *FBXW7*, and *SMAD4*, as well as activating mutations in *KRAS* (codon 12 and 146) and *PIK3CA* (codon 545 and 1047) were observed. Activating mutations in *BRAF* and *TGFBR1/2* mutations were observed in the hypermutated organoids, consistent with previous reports for primary CRC (Cancer Genome Atlas Network, 2012).

Mutations of genes in DNA mismatch repair (MMR)-associated pathways are associated with a hypermutated phenotype (Boland and Goel, 2010). Consistent with their classification as hypermutated CRC cases (Cancer Genome Atlas Network, 2012), missense mutations were present in *MSH3* in P7, and *POLE* mutations were detected in P10, P19a, and P19b. We did not observe mutations in MMR-associated genes in P24a and P24b and expression analysis showed normal levels of the pertinent genes. The culprit for hyper mutability thus remains to be identified for P24. The limited cohort size did not allow a statistical analysis for somatic copy number alterations to identify significant regions of amplification and deletions. However, manual inspection of the top regions identified by TCGA did reveal the presence of *ERBB2*-, *MYC*-, and *IGF2*-amplified organoids, as well as a reported gain of 13q in the non-hypermutated group (Figure 3C). In aggregate, these analyses demonstrate that organoid cultures faithfully capture the genomic features of the primary tumor from which they derive and much of the genomic diversity of CRC.

Most CRC cases carry activating mutations in the WNT pathway: inactivation mutations in *APC*, *FBXW7*, *AXIN2*, and *FAM123B*, or activating mutations in *CTNNB1* (Cancer Genome Atlas Network, 2012). Gene fusions involving the Wnt-agonistic *RSPO2* and *RSPO3* genes have been observed in 5%–10% of CRC (Seshagiri et al., 2012). *RNF43* encodes a negative regulator of the Wnt pathway, which serves to remove the Wnt receptor FZ in a negative feedback loop (Hao et al., 2012; Koo et al., 2012; de Lau et al., 2014). Recent sequencing efforts of gastric, ovary, and pancreatic neoplasias identified *RNF43* mutations (Jiao et al., 2014; Ryland et al., 2013; Wang et al., 2014), and *RNF43* mutations have been observed in

CRC (Giannakis et al., 2014; Ivanov et al., 2007; Koo et al., 2012). We found *APC* alterations in all but four of the organoids (P11, P19a/b, P28). Western blotting revealed P11 to express a truncated APC protein, pointing to a mutational event not covered by our exome-sequencing (Figure S3). The wtAPC organoid P28 carries an activating mutation in *CTNNB1* (T41A). In both P19a and P19b, we detected *RNF43* mutations: frameshifts at aa positions 659 and 355, respectively. Only the latter is predicted to affect protein function.

RNA Analysis of Normal and Tumor-Derived Organoids

Organoid cultures consist purely of epithelial cells. Therefore, the system allows for direct gene expression analysis without a contamination from mesenchyme, blood vessels, immune cells, etc. Normal colon-derived and tumor-derived organoids were plated under identical conditions in complete medium (+Wnt). After 3 days, RNA was analyzed using Affymetrix single transcript arrays. Figure 4A shows the correlation heatmap of the organoid samples. Normal colon-derived organoids clustered tightly together, while the tumor-derived organoids exhibited much more heterogeneity. Next, we searched for genes differentially expressed between normal and tumor organoids. Normal colon-derived organoids (Figure 4B) expressed genes of differentiated cells (e.g., the goblet cell markers *MUC1* and *MUC4* and the colonocyte marker *CA2*). Genes enriched in tumor organoids included cancer-associated genes such as *PROX1*, *BAMBI*, and *PTCH1* and the Wnt target gene *APCDD1* (Takahashi et al., 2002).

Several CRC classifications have been proposed based on RNA expression. We combined expression data from organoid samples and TCGA tissue samples and classified these in subtypes using the gene signatures by Sadanandam et al. (2013). Figure 4C displays the subtyping of the 22 organoid samples and 431 TCGA RNA sequencing (RNA-seq) tumor tissue samples. The heatmap shows the normalized scores of genes by samples, both sorted by subtype (see Experimental Procedures). Organoid samples were spread across the subtypes, with the transit-amplifying (TA) subtype being most frequently represented. The enterocyte subtype was not represented. In addition, the RNA expression data allowed expression analysis

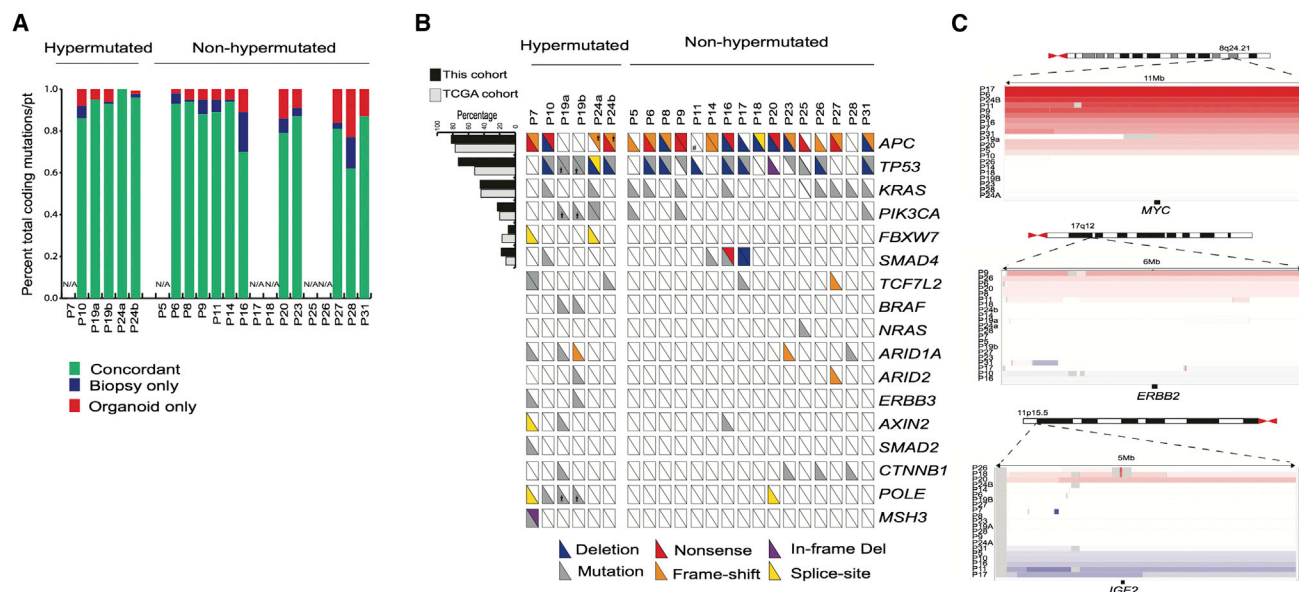


Figure 3. Genomic Alterations Found in CRC Are Represented in Organoid Cultures

(A) Concordance of somatic mutations detected in organoid and corresponding biopsies. Bar graph represents the proportion of coding alterations that are concordant between the biopsy and the corresponding organoid culture and those that are found only in organoid or biopsy specimen. N/A indicates cases in which exome-sequencing was not performed on the corresponding biopsy.

(B) Overview of the mutations found in the tumor organoids. The hash-mark in each box represents each allele and whether it was subject to deletion, mutation, frame-shift alteration, nonsense mutation or splice site mutation. Those alterations present in >10% of cases are compared to the percentage of cases reported by the TCGA CRC. *Indicates discordant mutations targeting the same gene between the two sites in P19 and P24. See also [Tables S11](#) and [S1J](#).

(C) Somatic copy-number alterations in organoids among commonly amplified genes identified in TCGA CRC.

See also [Figures S2](#) and [S3](#) and [Tables S1D–S1J](#).

of individual genes in organoids. *MLH1* expression was absent from two tumor organoids from patient 19 as well as from patient 7 (that is also mutant in *MSH3*) ([Figure S4](#)). In the two tumor organoids from P24, we did not detect expression changes in *MLH1* or any other MSI-associated gene.

Effect of Porcupine Inhibitor on RNF43 Mutant Organoids

Unlike most other WNT pathway mutations, *RNF43* mutations yield a cell that is hypersensitive to—yet still dependent on—secreted WNT. Array data confirmed the expression of several WNTs by the organoids ([Figure S5A](#)). The O-acyltransferase Porcupine is required for the secretion of WNTs and its inhibition prevents autocrine/paracrine activation of the pathway ([Kadowaki et al., 1996](#)). The small molecule porcupine inhibitor IWP2 ([Chen et al., 2009](#)) was tested on a small panel of the tumor organoids and strongly affected the *RNF43* mutant P19b organoid ([Figure 5A](#)). This observation implied that porcupine inhibition may be evaluated for treatment of the small subset of cancer patients mutant in *RNF43*.

Organoid Proof-of-Concept Drug Screen

Prompted by this, we developed a robotized drug sensitivity screen in 3D-organoid culture and correlated drug sensitivity with genomic features to identify molecular signatures associated with altered drug response. Organoid cultures were gently disrupted and plated on BME-coated 384-well plates in a 2%

BME solution. Organoids were left overnight before being drugged and left for 6 days before measuring cell number using CellTiter-Glo reagent. Drug sensitivity was represented by the half-maximal inhibitory concentration (IC_{50}), the slope of the dose-response curve, and area under the dose-response curve (AUC).

A bespoke 83 compound library was assembled for screening, including drugs in clinical use ($n = 25$), chemotherapeutics ($n = 10$), drugs previously investigated in or currently undergoing studies in clinical trials ($n = 29$), and experimental compounds to a diverse range of cancer targets ($n = 29$) ([Table S2A](#)). The library included the anti-EGFR antibody cetuximab, used clinically for *KRAS/NRAS/BRAF* wild-type CRC, as well as oxaliplatin and 5-FU, first line chemotherapeutics for CRC treatment. In total, 19 of 20 tumor organoids (from 18 different patients) were successfully screened in experimental triplicate, generating >5,000 measurements of organoid-drug interactions ([Table S2B](#)).

We incorporated a number of controls into the assay design. The median Z factor score, a measure of assay plate quality, across all screening plates was 0.62 ($n = 119$; upper and lower quartile = 0.85 and 0.3, respectively), consistent with an experimentally robust assay. We did observe some unexplained organoid-specific variation in assay plate quality. Dose-response measurements were performed in experimental triplicate or duplicate (on separate plates) and replicate AUC values were highly correlated (Pearson correlation [R_p] > 0.87) ([Figure 5B](#)).

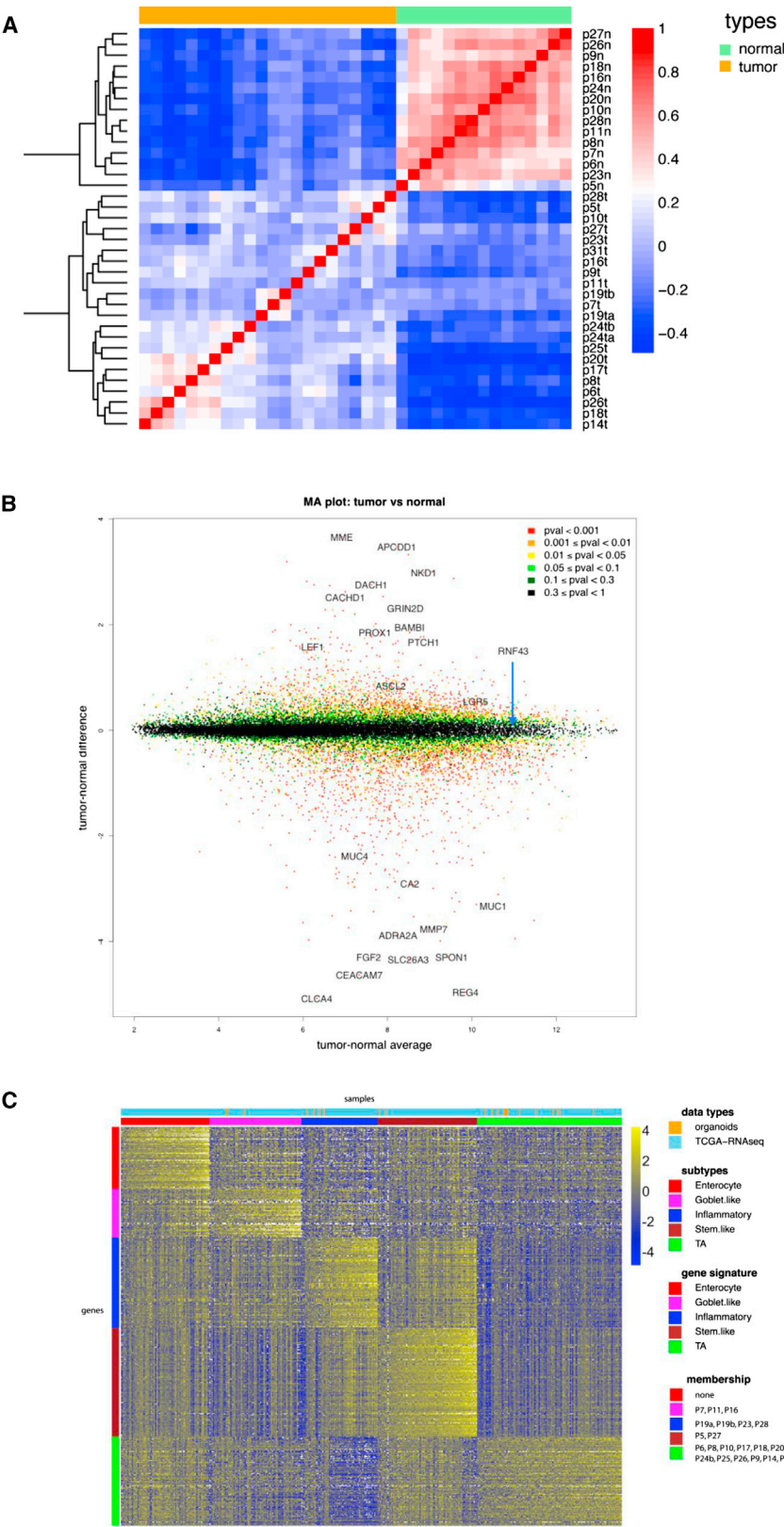


Figure 4. RNA Expression Analysis

(A) Correlation heat map of normal organoids versus tumor organoids based on 2,186 genes (the top 10% of genes in terms of SD). The normal organoids are very highly correlated with each other, whereas the tumor samples exhibit more heterogeneity. The colors represent pairwise Pearson correlations after the expression values have been logged and mean-centered for every gene. The hierarchical clustering is based on one minus correlation distance. The affix N = normal, T = tumor.

(B) MA plot of logged normal versus tumor gene expression. p values are computed with the R package limma, by comparing normal versus tumor gene expression. Cancer-associated genes (e.g., *APCDD1*, *PROX1*, and *PTCH1*) are shown in the top half.

(C) CRC molecular subtypes are represented by the organoid panel. Genes by samples heat map of normalized gene expression of 22 organoid samples and 431 TCGA RNA-seq tumor tissue samples, organized by subtype. Within each subtype, samples are sorted by their mean gene expression for the signature genes associated with that specific subtype.

See also [Figure S4](#).

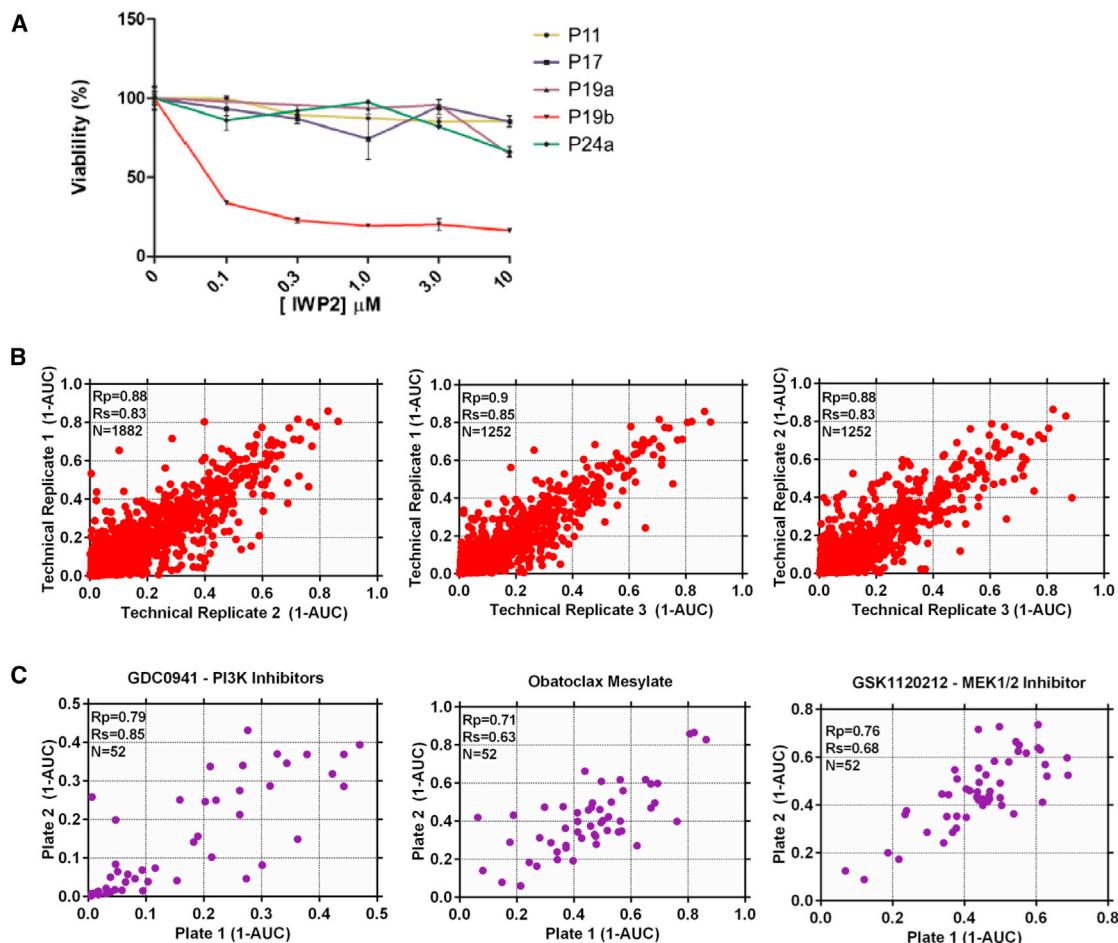


Figure 5. Development of a High-Throughput Drug Screening Assay Utilizing Organoid Models

(A) Autocrine/paracrine WNT signaling in P19b. A small panel of tumor organoids was incubated with increasing amounts of the Porcupine inhibitor IWP2. Growth of the *RNF43* mutant P19b was inhibited, indicative of dependency on autocrine/paracrine WNT signaling. Error bars indicate the SD of triplicate measurements. See also Figure S5.

(B) Scatterplot of (1-AUC) values for all technical replicates of drug screening data. Plots show the correlation between the three different technical replicates and each data point represents the (1-AUC) value for an individual organoid.

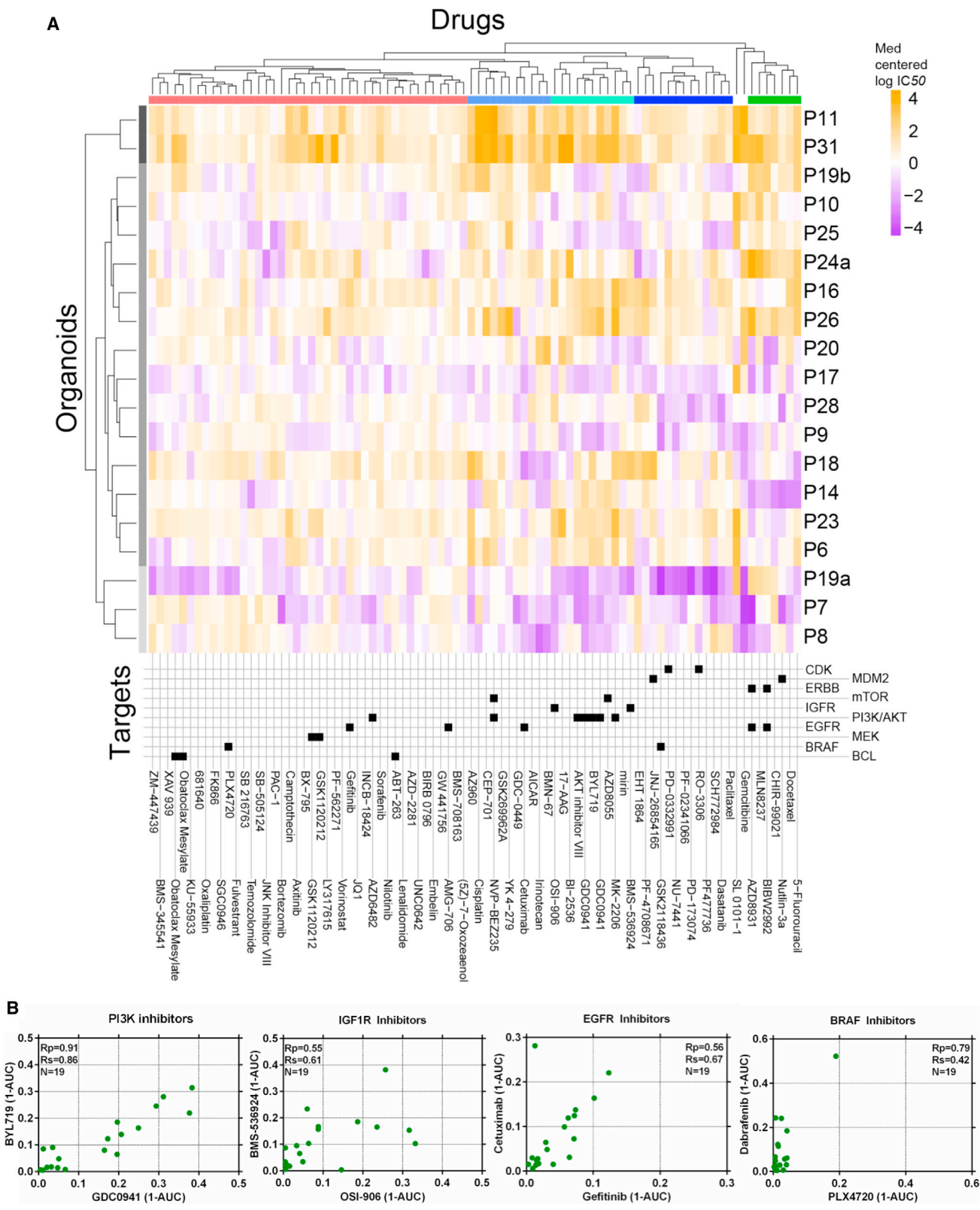
(C) Scatterplots of the correlation in (1-AUC) values for three compounds (GDC0941, obatoclox mesylate, and trametinib) screened twice during every screening run. Values are the mean of three technical replicates.

Furthermore, the compounds trametinib, GDC0941, and obatoclox mesylate were screened twice independently on separate assay plates and a good correlation was observed between the experimentally determined AUC values ($R_p = 0.79$, 0.71 , and 0.76 , respectively) (Figure 5C).

As a first validation, the only tumor organoid in the panel that was sensitive to the Porcupine inhibitor LGK974 was P19b (Figure S5B), confirming the observations made with IWP2 (Figure 5A). The clustering of compounds based on their IC_{50} values demonstrated a diverse range of sensitivities across the organoids and identified three major sub-groups (Figure 6A). One group was associated with sensitivity to a majority of the compounds (organoids P8, P7, and P19a), in contrast to the cluster (P31, P11) exhibiting insensitivity. The remaining organoids had intermediate sensitivity. Interestingly, the multifocal tumors P19a and P19b, derived from the same patient and

both carrying the BRAF V600E mutation, differed in their overall drug response profile. We observed clustering of drugs that inhibit the IGF1R and PI3K-AKT signaling pathways (Figure 6A), and compounds with similar nominal targets had comparable activity across the organoid collection. For example, a similar sensitivity pattern was observed for the PI3K inhibitors GDC0941 and BYL719 (α -selective), the IGF1R inhibitors OSI-906 and BMS-536924, EGFR inhibitors cetuximab and gefitinib, and the BRAF inhibitors dabrafenib and PLX4720 (Figure 6B). All but one of the organoids displayed a lack of sensitivity to BRAF inhibition. P19a, a BRAF V600E mutant organoid, displayed partial sensitivity to dabrafenib with an IC_{50} of $0.5 \mu M$, comparable to IC_{50} values of BRAF V600E colorectal cancer cell lines (range 0.004 – $2.55 \mu M$; average $0.96 \mu M$).

To identify genetic correlates between individual oncogenic mutations and drug response, we performed a multivariate



(legend on next page)

analysis of variance (MANOVA) incorporating IC_{50} values and slopes of the corresponding dose-response curves, with MSI-status as a covariate. Complete drug sensitivity and genomic data sets were available for 18 organoids and used for this analysis. The analysis included 16 genes identified as mutated, amplified, or deleted in CRC (referred to as mutant genes) as described by Lawrence et al. (2014) (Table S3). The MANOVA identified a subset (12 of 864, ~1%) of gene-drug associations as statistically significant ($p < 0.005$, incorporating a 30% false discovery rate [FDR]) (Table S4). These results were further filtered based on the magnitude of the effect size on the IC_{50} values of wild-type versus mutant cell line populations (effect size >2 ; Cohen's D), and correlations identified due to a singlet outlier organoids were removed. This resulted in the identification of one high confidence gene-drug association already reported in the literature (Vassilev et al., 2004). Loss-of-function mutations of the tumor suppressor TP53 were associated with resistance to nutlin-3a ($p = 0.0018$), an inhibitor of MDM2 (Figure 7A). Of the four organoids that were wild-type for TP53 by DNA sequencing, only P18 was (unexpectedly) insensitive to nutlin-3a. However, immunohistochemistry of p53 in P18 revealed the protein to be stabilized, indicative of functional inactivation of the p53 pathway (Figure 7B).

We could also readily detect resistance to the anti-EGFR inhibitors cetuximab and BIBW2992 (afatinib) in the setting of KRAS mutant organoids ($p = 0.008$ /FDR 37% and $p = 0.029$ /FDR 54%, respectively), although these associations were below statistical significance when considering an FDR $<30\%$ (Figures 7C and S6). Of the KRAS wild-type organoids, a subset 2/10 was insensitive to cetuximab, including P19b that has a BRAF mutation, a known mediator of cetuximab resistance (Di Nicolantonio et al., 2008). For the remaining organoid, further mechanisms beyond mutated KRAS/NRAS/BRAF are likely to be involved in cetuximab resistance (De Roock et al., 2010; Vecchione, 2014).

We also identified a number of compounds with differential activity in the absence of an apparent genetic biomarker (Figure 7D). For example, a subset of organoids was exquisitely sensitive to the AKT1/2 inhibitor MK2206. Similarly, we observed distinct subsets of organoids that are exquisitely sensitive to the pan-ERBB inhibitor AZD8931 and the chemotherapeutic gemcitabine. We also performed a validation screen with 11 of the original 83 compounds across the organoid panel and compared the measured responses (Figure S7; Table S5). We observed positive correlation for all compounds and nine exhibited good to fair reproducibility as indicated by an R_p of 0.5 or greater (Figures 7E and 7F). Variation within the assay was likely due to inherent technical noise, biological variation, and sensitivity to outlier data points due to the small number of organoids.

In summary, the successful application of organoids in a systematic and unbiased high-throughput drug screen to

identify clinically relevant biomarkers demonstrates the feasibility and utility of organoid technology for investigating the molecular basis of drug response. Furthermore, the identification of putative novel molecular markers has opened avenues for further investigation of drug sensitivity in CRC. The current analysis is still constrained by the relatively small number of patients. The derivation of a significantly larger organoid collection would increase the representation of rare genotypes and the statistical power to detect molecular markers of drug response.

DISCUSSION

Cancer cell lines have served for many years as the workhorse model in cancer research. Recent studies have exploited high-throughput screening of large panels of cancer cell lines to identify drug-sensitivity patterns and to correlate drug sensitivity to genomic alterations (Barretina et al., 2012; Garnett et al., 2012). From these high-throughput cell-line-based studies, a picture emerges of a complex network of biological factors that affect sensitivity to the majority of cancer drugs. For instance, no direct relationship may exist between sensitivity to a certain drug and a single genomic alteration. Instead, difficult-to-find, complex interactions between multiple genomic alterations may determine drug sensitivity outcome. Thus, with currently available insights, it remains a challenge to develop algorithms that accurately predict the drug sensitivity of a patient's tumor based on the spectrum of genomic alterations present, in the context of the unique genetic background.

Two approaches to determine directly the drug sensitivity in a patient-derived sample have been quite widely exploited, namely the short-term culture of tumor sections (Centenera et al., 2013), and xeno-transplantation of the tumor into immunodeficient mice (Jin et al., 2010; Tentler et al., 2012). Short-term culture allows for in vitro screening at a reasonably large scale, but is constrained by the limited proliferative capacity of the cultures. Xenotransplantation allows for in vivo screening but is resource-intensive due to the need for large mouse colonies. It thus appears of interest to develop additional technologies that allow the combination of sequencing and high-throughput drug screening in patient-derived samples. Here, we demonstrate that the organoid culture platform can be exploited for genomic and functional studies at the level of the individual patient at a scale that cannot be achieved by existing approaches. Our organoid drug screening assay generates reproducible high quality drug sensitivity data, positive correlation of biological replicates, and reproducible activity of compounds inhibiting the same target. By connecting genetic and drug sensitivity data, we were able to confirm the activity of cetuximab in a subset of KRAS wild-type organoids reflecting observations made in the clinic (De Roock et al., 2010) as well as Nutlin-3a effectiveness in TP53 wild-type organoids. Furthermore, we describe

Figure 6. Heatmap of IC_{50} s of All 85 Compounds against 19 Colorectal Cancer Organoids

(A) Organoids have been clustered based on their IC_{50} values across the drug panel. The drug names and their nominal target(s) are provided in the bottom panel. (B) Drugs with the same nominal targets have similar activity profiles across the organoid panel. (1-AUC) values are plotted for inhibitor of PI3K (GDC0941 and BYL719), IGF1R (OSI-906 and BMS-536924), EGFR (cetuximab and gefitinib), and BRAF (PLX4720 and dabrafenib).

See also Tables S2A and S2B.

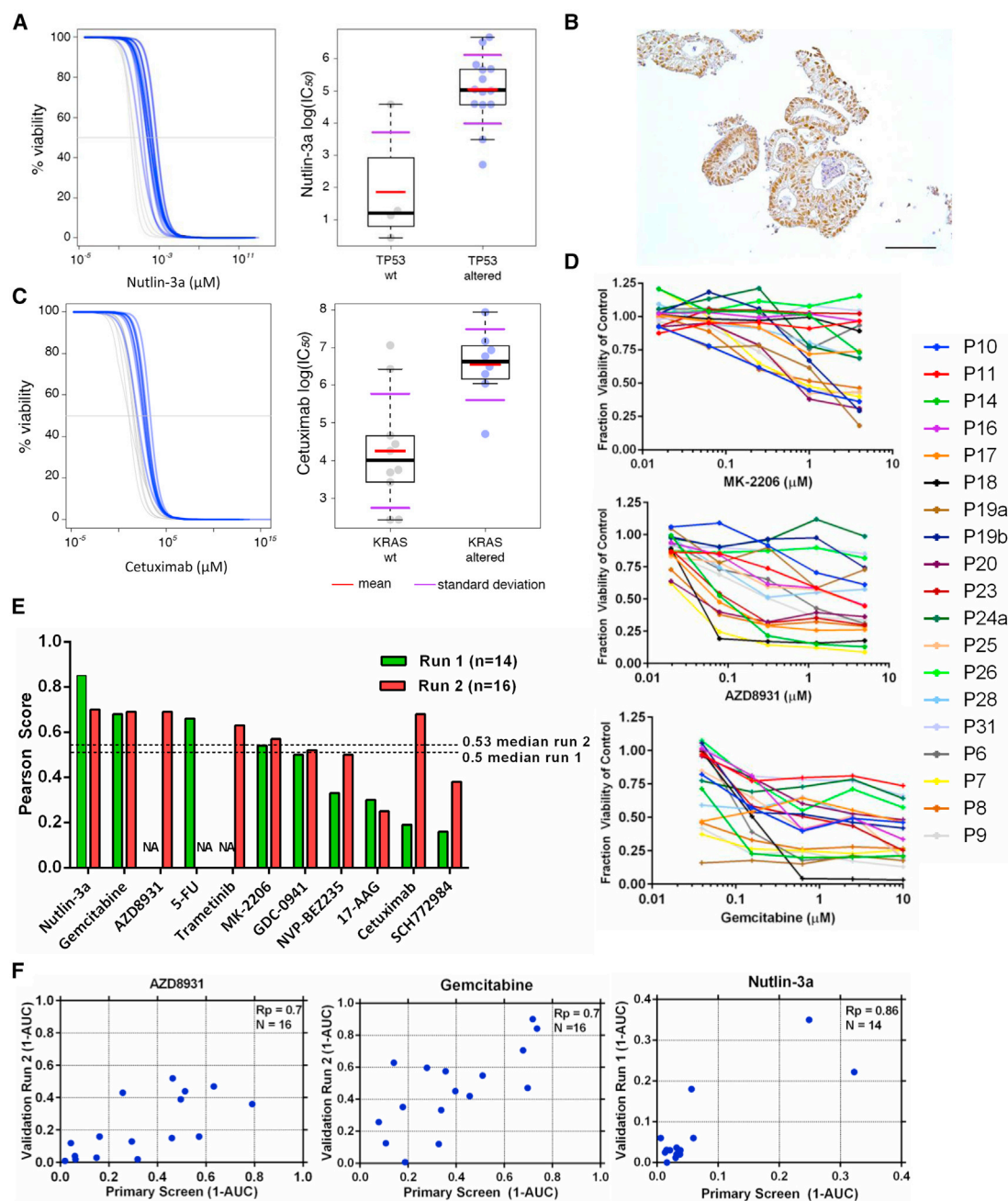


Figure 7. Gene-Drug Associations and Differential Drug Sensitivity Profiles of Interest

(A) Association of TP53 mutational status with nutlin-3a response. Viability response curves of the altered (blue) and wild-type organoids (gray) as well as scatter plots of cell line IC_{50} (μM) values are shown. IC_{50} values are on a natural logarithmic scale. Each circle represents one cell line, red bars indicate geometric means of IC_{50} values and black bold bars indicate median $\log IC_{50}$ values. Box top/low bounds indicate upper/lower quartiles, and whiskers (indicated by the dashed lines) extend to extreme values (minimal and maximal) excluding outliers (i.e., whose value is more than 3/2 times the upper quartile and less than 3/2 times the lower quartile). Purple bar positions on the y axis indicate means $\pm \log IC_{50}$ SD.

(B) Immunohistochemical staining showing stabilization of TP53 in organoid P18. Scale bar, 100 μM .

(C) Association of KRAS status and cetuximab response. Colors and symbols coding is the same as (A).

(D) Dose-response curves after 6 days treatment with MK2206, AZD8931, and gemcitabine.

(E) Reproducibility of drug response profiles for 11 drugs. The Pearson correlation score of (1-AUC) values from the primary screen compared to (1-AUC) values from validation screens are used for comparison. The validation screen was performed twice (run 1 and 2) with >1 month elapsed between each screen. NA, data unavailable for this drug.

(F) The correlation of 1-AUC values from the primary and validation screens for AZD8931, gemcitabine, and nutlin-3a.

See also Figures S6 and S7 and Tables S3, S4, and S5.

the differential activity of a handful of clinical and preclinical compounds (gemcitabine, MK2206, and AZD8941).

Tumors are composed of a mixture of sub-clones that coevolve through a Darwinian selection process. This cellular heterogeneity and phenotypic variation allows the emergence of a complex clonal architecture, which underpins important features such as drug resistance and metastatic potential (Burrell et al., 2013). Our CCF analysis of clonal structure determined that almost all of the biopsies were polyclonal at the time of resection, and this is reflected to varying extent in the corresponding organoid culture. The ability to capture sub-clonal populations in *in vitro* organoid culture should enable more predictive modeling of patient responses to therapy. In many respects, the clonal selection and heterogeneity observed in organoids is similar to PDX models of cancers (Eirew et al., 2015). For both models, understanding the factors that affect tumor heterogeneity and evolution, and how heterogeneity impacts on drug response, will be important to fully exploit their potential for predicting patient responses.

We perceive patient-derived organoids to be used to directly test drug sensitivity of the tumor in a personalized treatment approach. For this, we envision organoids to be tested against a limited number of clinically approved drugs within weeks after derivation. While building this pilot biobank, we observed that normal epithelial tissue always yield good numbers of organoids within weeks, while significant differences in “take rates” were observed between patients’ tumor organoids. Crucial for this approach to be effective, is to decrease the time needed to derive and expand the organoids. In conclusion, tumor organoids may fill the gap between cancer genetics and patient trials, complement cell-line- and xenograft-based drug studies, and allow personalized therapy design.

EXPERIMENTAL PROCEDURES

Human Tissues

Colonic tissues were obtained from The Diaconessen Hospital Utrecht with informed consent and the study was approved by the ethical committee. All patients were diagnosed with colorectal cancer. From the resected colon segment, normal as well as tumor tissue was isolated. The isolation of healthy crypts and tumor epithelium was performed essentially as described by Sato et al. (2011).

Organoid Culture

Healthy tissue-derived organoids were cultured in Human Intestinal Stem Cell medium (HISC). The composition of HISC is: Basal culture medium with 50% Wnt conditioned medium, 20% R-Spondin conditioned medium, 10% Noggin conditioned medium, 1 × B27, 1.25 mM n-Acetyl Cysteine, 10 mM Nicotinamide, 50 ng/ml human EGF, 10 nM Gastrin, 500 nM A83-01, 3 μM SB202190, 10 nM Prostaglandin E2, and 100 μg/ml Primocin (Vivogen). Tumor organoids were cultured in HISC minus Wnt. See the [Extended Experimental Procedures](#) for a detailed description.

Whole-Exome Sequencing and Copy-Number Analysis

For each sample, ~250 ng of DNA was sheared and subject to whole-exome sequencing using the Agilent v2 capture probe set and sequenced by HiSeq2500 using 76 base pair reads, as previously described (Fisher et al., 2011; Imielinski et al., 2012). A median 9.6 Gb of unique sequence was generated for each sample (Table S1A).

Sequence data were locally realigned to improve sensitivity and reduce alignment artifacts prior to identification of mutations, insertions, and deletions

as previously described (Cibulskis et al., 2013; DePristo et al., 2011; Ojesina et al., 2014).

Somatic copy-number analysis was performed using segmented copy-number profiles generated from whole-exome sequencing using the SegSeq algorithm (Table S1D) (Chiang et al., 2009). The procedure is described in detail in the [Extended Experimental Procedures](#).

Organoid Data Processing

RNA from 22 organoid tumor samples and 15 paired normal samples was hybridized on Affymetrix Human Gene 2.0 ST arrays. The raw CEL files were processed with Affymetrix Power Tools using the Hg19 genome build and NetAffx annotation dating from 09-30-2012. Between-array normalization was performed using rma-sketch, within APT. This resulted in an intensity matrix of 21,681 genes by 37 samples. For analysis of individual genes, data were analyzed using the R2 web application, which is freely available at <http://r2.amc.nl>.

To subtype the samples, we used the gene signature published by Sadanandam et al. (2013). The procedure is described in detail in the [Extended Experimental Procedures](#).

Organoid Viability Assays

Eight microliters of ~7 mg/ml BME was dispensed in to 384-well microplates and allowed to polymerize. Organoids were mechanically dissociated by pipetting before being resuspended in 2% BME/growth media (15–20,000 organoids/ml) and dispensed into drug wells. The following day a 5-point 4-fold dilution series of each compound was dispensed using liquid handling robotics and cell viability assayed using CellTiter-Glo (Promega) following 6 days of drug incubation. All screening plates were subjected to stringent quality control measures and a Z factor score comparing negative and positive control wells calculated. Dose-response curves were fitted to the luminescent signal intensities utilizing a method previously described (Garrett et al., 2012). Further information of the compounds used, data-fitting algorithm, and validation screen can be found in the [Extended Experimental Procedures](#).

Systematic Multivariate Analysis of Variance

We excluded from the analysis drugs with no IC₅₀ values falling within the range of tested concentrations. For each of the remaining drugs, we assembled an 18 × 2 matrix Y composed by two vectors of length n = 18, containing IC₅₀ values and dose-response curve slopes β, respectively, obtained by treating 18 organoids with the drug under consideration. A multivariate analysis of variance (MANOVA) model was then fitted to this drug response data matrix with factors including the microsatellite stability status of the organoids and the status (altered or wild-type) of 16 genomic features ([Extended Experimental Procedures](#)). Significance and effect size scores were obtained for each of the genomic-feature/drug pairs. Q values were subsequently obtained by correcting the MANOVA p values for multiple hypotheses testing, and a threshold of 30% of positive false discovery rate, IC₅₀, and effect size >2 (as quantified by the Cohen's D) was used to identify significant associations.

ACCESSION NUMBERS

The accession number for the healthy and tumor organoid array data reported in this paper is GEO: GSE64392. The accession number for the single organoid RNA-seq data is GEO: GSE65253.

SUPPLEMENTAL INFORMATION

Supplemental Information includes [Extended Experimental Procedures](#), seven figures, five tables, and one data file and can be found with this article online at <http://dx.doi.org/10.1016/j.cell.2015.03.053>.

AUTHOR CONTRIBUTIONS

M.v.d.W. derived, maintained, and analyzed organoid cultures. H.E.F. developed, performed, and analyzed the organoid drug screen. J.M.F. analyzed

sequencing data. G.B. analyzed RNA expression data. R.G.J.V. organized ethical approval. F.I. performed analyses and statistical inferences on the drug screening data supervised by J.S.R. A.P. and W.v.H. performed surgery. J.v.G. isolated tumor and normal tissue from resected material. A.T.M. performed cancer cell fraction analysis. L.K. performed single organoid transcriptomics supervised by A.v.O. A.M. assisted in drug screening. J.B. performed immunostainings and assisted in culturing organoids together with S.J. and S.B. P.v.d.S. and R.V. processed and analyzed RNA, supervised by R.V. V.S.W.L. performed APC western analysis. S.S. processed DNA samples for exome sequencing. C.S.P., K.C., S.L.C., A.M., M.S.L., L.L., and C.S. helped in processing and analyzing sequencing data. G.G. and M.M. supervised the sequencing and analysis. L.W. supervised the RNA analysis. M.R.S., U.M., M.G., and H.C. participated in the development of the project concept. M.G. aided in drug data analysis. M.v.d.W. and H.E.F. participated in data analysis and project design. M.v.d.W., H.E.F., J.M.F., M.M., M.J.G., and H.C. wrote the manuscript.

ACKNOWLEDGMENTS

Thanks to the Broad Institute Genomics Platform for processing and generating the sequence data, the cell line screening, and analysis team at the Sanger Institute, Liliene Wijnaendts and Joost Oudejans for help with tumor dissections, Harry Begthel and Jeroen Korving for histology, Sepideh Darakhshan for help with organoids, and the members of the M.M., M.J.G., and H.C. laboratories for support. M.v.d.W. is supported by Stichting Virtutis Opus and Stichting Vrienden van het Hubrecht. R.G.J.V., S.J., and J.B. are supported by a Grant from Alpe d'Huizes/KWF. This work was supported with a grant from the Dutch Cancer Society to M.S. (H1/2014-6919). The organoids are available for academic research upon evaluation of a research proposal by the Medical Ethical Comity. M.M. received a commercial research grant from Bayer and has ownership interest (including patents) in and is a consultant/advisory board member for Foundation Medicine.

Received: October 29, 2014

Revised: January 20, 2015

Accepted: March 18, 2015

Published: May 7, 2015

REFERENCES

- Barker, N., van Es, J.H., Kuipers, J., Kujala, P., van den Born, M., Cozijnsen, M., Haegebarth, A., Korving, J., Begthel, H., Peters, P.J., and Clevers, H. (2007). Identification of stem cells in small intestine and colon by marker gene *Lgr5*. *Nature* 449, 1003–1007.
- Barretina, J., Caponigro, G., Stransky, N., Venkatesan, K., Margolin, A.A., Kim, S., Wilson, C.J., Lehár, J., Kryukov, G.V., Sonkin, D., et al. (2012). The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. *Nature* 483, 603–607.
- Bass, A.J., Lawrence, M.S., Bracci, L.E., Ramos, A.H., Drier, Y., Cibulskis, K., Sougnez, C., Voet, D., Saksena, G., Sivachenko, A., et al. (2011). Genomic sequencing of colorectal adenocarcinomas identifies a recurrent *VT11A-TCF7L2* fusion. *Nat. Genet.* 43, 964–968.
- Boland, C.R., and Goel, A. (2010). Microsatellite instability in colorectal cancer. *Gastroenterology* 138, 2073–2087.
- Burrell, R.A., McGranahan, N., Bartek, J., and Swanton, C. (2013). The causes and consequences of genetic heterogeneity in cancer evolution. *Nature* 501, 338–345.
- Cancer Genome Atlas Network (2012). Comprehensive molecular characterization of human colon and rectal cancer. *Nature* 487, 330–337.
- Carmon, K.S., Gong, X., Lin, Q., Thomas, A., and Liu, Q. (2011). R-spondins function as ligands of the orphan receptors *LGR4* and *LGR5* to regulate *Wnt*/beta-catenin signaling. *Proc. Natl. Acad. Sci. USA* 108, 11452–11457.
- Carter, S.L., Cibulskis, K., Helman, E., McKenna, A., Shen, H., Zack, T., Laird, P.W., Onofrio, R.C., Winckler, W., Weir, B.A., et al. (2012). Absolute quantification of somatic DNA alterations in human cancer. *Nat. Biotechnol.* 30, 413–421.
- Centenera, M.M., Raj, G.V., Knudsen, K.E., Tilley, W.D., and Butler, L.M. (2013). Ex vivo culture of human prostate tissue and drug development. *Nat. Rev. Urol.* 10, 483–487.
- Chen, B., Dodge, M.E., Tang, W., Lu, J., Ma, Z., Fan, C.-W., Wei, S., Hao, W., Kilgore, J., Williams, N.S., et al. (2009). Small molecule-mediated disruption of *Wnt*-dependent signaling in tissue regeneration and cancer. *Nat. Chem. Biol.* 5, 100–107.
- Chiang, D.Y., Getz, G., Jaffe, D.B., O'Kelly, M.J.T., Zhao, X., Carter, S.L., Russ, C., Nusbaum, C., Meyerson, M., and Lander, E.S. (2009). High-resolution mapping of copy-number alterations with massively parallel sequencing. *Nat. Methods* 6, 99–103.
- Cibulskis, K., Lawrence, M.S., Carter, S.L., Sivachenko, A., Jaffe, D., Sougnez, C., Gabriel, S., Meyerson, M., Lander, E.S., and Getz, G. (2013). Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. *Nat. Biotechnol.* 31, 213–219.
- de Lau, W., Barker, N., Low, T.Y., Koo, B.-K., Li, V.S.W., Teunissen, H., Kujala, P., Haegebarth, A., Peters, P.J., van de Wetering, M., et al. (2011). *Lgr5* homologues associate with *Wnt* receptors and mediate R-spondin signalling. *Nature* 476, 293–297.
- de Lau, W., Peng, W.C., Gros, P., and Clevers, H. (2014). The R-spondin/*Lgr5*/*Rnf43* module: regulator of *Wnt* signal strength. *Genes Dev.* 28, 305–316.
- De Rooij, W., Claes, B., Bernasconi, D., De Schutter, J., Biesmans, B., Fountzilas, G., Kalogeris, K.T., Kotoula, V., Papamichael, D., Laurent-Puig, P., et al. (2010). Effects of *KRAS*, *BRAF*, *NRAS*, and *PIK3CA* mutations on the efficacy of cetuximab plus chemotherapy in chemotherapy-refractory metastatic colorectal cancer: a retrospective consortium analysis. *Lancet Oncol.* 11, 753–762.
- DePristo, M.A., Banks, E., Poplin, R., Garimella, K.V., Maguire, J.R., Hartl, C., Philippakis, A.A., del Angel, G., Rivas, M.A., Hanna, M., et al. (2011). A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat. Genet.* 43, 491–498.
- Di Nicolantonio, F., Martini, M., Molinari, F., Sartore-Bianchi, A., Arena, S., Saletti, P., De Dosso, S., Mazzucchelli, L., Frattini, M., Siena, S., and Bardelli, A. (2008). Wild-type *BRAF* is required for response to panitumumab or cetuximab in metastatic colorectal cancer. *J. Clin. Oncol.* 26, 5705–5712.
- Eirew, P., Steif, A., Khattra, J., Ha, G., Yap, D., Farahani, H., Gelmon, K., Chia, S., Mar, C., Wan, A., et al. (2015). Dynamics of genomic clones in breast cancer patient xenografts at single-cell resolution. *Nature* 518, 422–426.
- Fearon, E.R. (2011). Molecular genetics of colorectal cancer. *Annu. Rev. Pathol.* 6, 479–507.
- Fearon, E.R., and Vogelstein, B. (1990). A genetic model for colorectal tumorigenesis. *Cell* 61, 759–767.
- Fisher, S., Barry, A., Abreu, J., Minie, B., Nolan, J., Delorey, T.M., Young, G., Fennell, T.J., Allen, A., Ambrogio, L., et al. (2011). A scalable, fully automated process for construction of sequence-ready human exome targeted capture libraries. *Genome Biol.* 12, R1.
- Futreal, P.A., Coin, L., Marshall, M., Down, T., Hubbard, T., Wooster, R., Rahman, N., and Stratton, M.R. (2004). A census of human cancer genes. *Nat. Rev. Cancer* 4, 177–183.
- Garnett, M.J., Edelman, E.J., Heidorn, S.J., Greenman, C.D., Dastur, A., Lau, K.W., Greninger, P., Thompson, I.R., Luo, X., Soares, J., et al. (2012). Systematic identification of genomic markers of drug sensitivity in cancer cells. *Nature* 483, 570–575.
- Garraway, L.A., and Lander, E.S. (2013). Lessons from the cancer genome. *Cell* 153, 17–37.
- Giannakis, M., Hodis, E., Jasmine Mu, X., Yamauchi, M., Rosenbluh, J., Cibulskis, K., Saksena, G., Lawrence, M.S., Qian, Z.R., Nishihara, R., et al. (2014). *RNF43* is frequently mutated in colorectal and endometrial cancers. *Nat. Genet.* 46, 1264–1266.

- Hao, H.-X., Xie, Y., Zhang, Y., Charlat, O., Oster, E., Avello, M., Lei, H., Mickanin, C., Liu, D., Ruffner, H., et al. (2012). ZNRF3 promotes Wnt receptor turnover in an R-spondin-sensitive manner. *Nature* 485, 195–200.
- Haramis, A.-P.G., Begthel, H., van den Born, M., van Es, J., Jonkhoe, S., Offerhaus, G.J.A., and Clevers, H. (2004). De novo crypt formation and juvenile polyposis on BMP inhibition in mouse intestine. *Science* 303, 1684–1686.
- Imielinski, M., Berger, A.H., Hammerman, P.S., Hernandez, B., Pugh, T.J., Hodis, E., Cho, J., Suh, J., Capelletti, M., Sivachenko, A., et al. (2012). Mapping the hallmarks of lung adenocarcinoma with massively parallel sequencing. *Cell* 150, 1107–1120.
- Ivanov, I., Lo, K.C., Hawthorn, L., Cowell, J.K., and Ionov, Y. (2007). Identifying candidate colon cancer tumor suppressor genes using inhibition of nonsense-mediated mRNA decay in colon cancer cells. *Oncogene* 26, 2873–2884.
- Jiao, Y., Yonescu, R., Offerhaus, G.J.A., Klimstra, D.S., Maitra, A., Eshleman, J.R., Herman, J.G., Poh, W., Pelosof, L., Wolfgang, C.L., et al. (2014). Whole-exome sequencing of pancreatic neoplasms with acinar differentiation. *J. Pathol.* 232, 428–435.
- Jin, K., Teng, L., Shen, Y., He, K., Xu, Z., and Li, G. (2010). Patient-derived human tumour tissue xenografts in immunodeficient mice: a systematic review. *Clin. Transl. Oncol.* 12, 473–480.
- Kadowaki, T., Wilder, E., Klingensmith, J., Zachary, K., and Perrimon, N. (1996). The segment polarity gene *porcupine* encodes a putative multitransmembrane protein involved in Wingless processing. *Genes Dev.* 10, 3116–3128.
- Kim, K.-A., Kakitani, M., Zhao, J., Oshima, T., Tang, T., Binnerts, M., Liu, Y., Boyle, B., Park, E., Emtage, P., et al. (2005). Mitogenic influence of human R-spondin1 on the intestinal epithelium. *Science* 309, 1256–1259.
- Koo, B.-K., Spit, M., Jordens, I., Low, T.Y., Stange, D.E., van de Wetering, M., van Es, J.H., Mohammed, S., Heck, A.J.R., Maurice, M.M., and Clevers, H. (2012). Tumour suppressor RNF43 is a stem-cell E3 ligase that induces endocytosis of Wnt receptors. *Nature* 488, 665–669.
- Korinek, V., Barker, N., Moerer, P., van Donselaar, E., Huls, G., Peters, P.J., and Clevers, H. (1998). Depletion of epithelial stem-cell compartments in the small intestine of mice lacking Tcf-4. *Nat. Genet.* 19, 379–383.
- Kuhnert, F., Davis, C.R., Wang, H.-T., Chu, P., Lee, M., Yuan, J., Nusse, R., and Kuo, C.J. (2004). Essential requirement for Wnt signaling in proliferation of adult small intestine and colon revealed by adenoviral expression of Dickkopf-1. *Proc. Natl. Acad. Sci. USA* 101, 266–271.
- Landau, D.A., Carter, S.L., Stojanov, P., McKenna, A., Stevenson, K., Lawrence, M.S., Sougnez, C., Stewart, C., Sivachenko, A., Wang, L., et al. (2013). Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. *Cell* 152, 714–726.
- Lawrence, M.S., Stojanov, P., Mermel, C.H., Robinson, J.T., Garraway, L.A., Golub, T.R., Meyerson, M., Gabriel, S.B., Lander, E.S., and Getz, G. (2014). Discovery and saturation analysis of cancer genes across 21 tumour types. *Nature* 505, 495–501.
- Lengauer, C., Kinzler, K.W., and Vogelstein, B. (1997). Genetic instability in colorectal cancers. *Nature* 386, 623–627.
- Liu, X., Ory, V., Chapman, S., Yuan, H., Albanese, C., Kallakury, B., Timofeeva, O.A., Nealon, C., Dakic, A., Simic, V., et al. (2012). ROCK inhibitor and feeder cells induce the conditional reprogramming of epithelial cells. *Am. J. Pathol.* 180, 599–607.
- Ojesina, A.I., Lichtenstein, L., Freeman, S.S., Pedamallu, C.S., Imaz-Rosshandler, I., Pugh, T.J., Cherniack, A.D., Ambrogio, L., Cibulskis, K., Bertelsen, B., et al. (2014). Landscape of genomic alterations in cervical carcinomas. *Nature* 506, 371–375.
- Pinto, D., Gregorieff, A., Begthel, H., and Clevers, H. (2003). Canonical Wnt signals are essential for homeostasis of the intestinal epithelium. *Genes Dev.* 17, 1709–1713.
- Ryland, G.L., Hunter, S.M., Doyle, M.A., Rowley, S.M., Christie, M., Allan, P.E., Bowtell, D.D.L., Gorringe, K.L., and Campbell, I.G.; Australian Ovarian Cancer Study Group (2013). RNF43 is a tumour suppressor gene mutated in mucinous tumours of the ovary. *J. Pathol.* 229, 469–476.
- Sadanandam, A., Lyssiotis, C.A., Homiczko, K., Collisson, E.A., Gibb, W.J., Wullschlegel, S., Ostos, L.C.G., Lannon, W.A., Grotzinger, C., Del Rio, M., et al. (2013). A colorectal cancer classification system that associates cellular phenotype and responses to therapy. *Nat. Med.* 19, 619–625.
- Sato, T., Vries, R.G., Snippert, H.J., van de Wetering, M., Barker, N., Stange, D.E., van Es, J.H., Abo, A., Kujala, P., Peters, P.J., and Clevers, H. (2009). Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. *Nature* 459, 262–265.
- Sato, T., Stange, D.E., Ferrante, M., Vries, R.G.J., Van Es, J.H., Van den Brink, S., Van Houdt, W.J., Pronk, A., Van Gorp, J., Siersema, P.D., and Clevers, H. (2011). Long-term expansion of epithelial organoids from human colon, adenoma, adenocarcinoma, and Barrett's epithelium. *Gastroenterology* 141, 1762–1772.
- Seshagiri, S., Stawiski, E.W., Durinck, S., Modrusan, Z., Storm, E.E., Conboy, C.B., Chaudhuri, S., Guan, Y., Janakiraman, V., Jaiswal, B.S., et al. (2012). Recurrent R-spondin fusions in colon cancer. *Nature* 488, 660–664.
- Takahashi, M., Fujita, M., Furukawa, Y., Hamamoto, R., Shimokawa, T., Miwa, N., Ogawa, M., and Nakamura, Y. (2002). Isolation of a novel human gene, APCDD1, as a direct target of the beta-Catenin/T-cell factor 4 complex with probable involvement in colorectal carcinogenesis. *Cancer Res.* 62, 5651–5656.
- Tentler, J.J., Tan, A.C., Weekes, C.D., Jimeno, A., Leong, S., Pitts, T.M., Arcaroli, J.J., Messersmith, W.A., and Eckhardt, S.G. (2012). Patient-derived tumour xenografts as models for oncology drug development. *Nat. Rev. Clin. Oncol.* 9, 338–350.
- Vassilev, L.T., Vu, B.T., Graves, B., Carvajal, D., Podlaski, F., Filipovic, Z., Kong, N., Kammlott, U., Lukacs, C., Klein, C., et al. (2004). In vivo activation of the p53 pathway by small-molecule antagonists of MDM2. *Science* 303, 844–848.
- Vecchione, L. (2014). Optimization of anti-EGFR treatment of advanced colorectal cancer. *Curr. Colorectal Cancer Rep.* 10, 263–271.
- Vogelstein, B., Papadopoulos, N., Velculescu, V.E., Zhou, S., Diaz, L.A., Jr., and Kinzler, K.W. (2013). Cancer genome landscapes. *Science* 339, 1546–1558.
- Wang, K., Yuen, S.T., Xu, J., Lee, S.P., Yan, H.H.N., Shi, S.T., Siu, H.C., Deng, S., Chu, K.M., Law, S., et al. (2014). Whole-genome sequencing and comprehensive molecular profiling identify new driver mutations in gastric cancer. *Nat. Genet.* 46, 573–582.
- Wong, V.W.Y., Stange, D.E., Page, M.E., Buczak, S., Wabik, A., Itami, S., van de Wetering, M., Poulosom, R., Wright, N.A., Trotter, M.W.B., et al. (2012). Lrig1 controls intestinal stem-cell homeostasis by negative regulation of ErbB signalling. *Nat. Cell Biol.* 14, 401–408.

Retraction Notice to: A Self-Produced Trigger for Biofilm Disassembly that Targets Exopolysaccharide

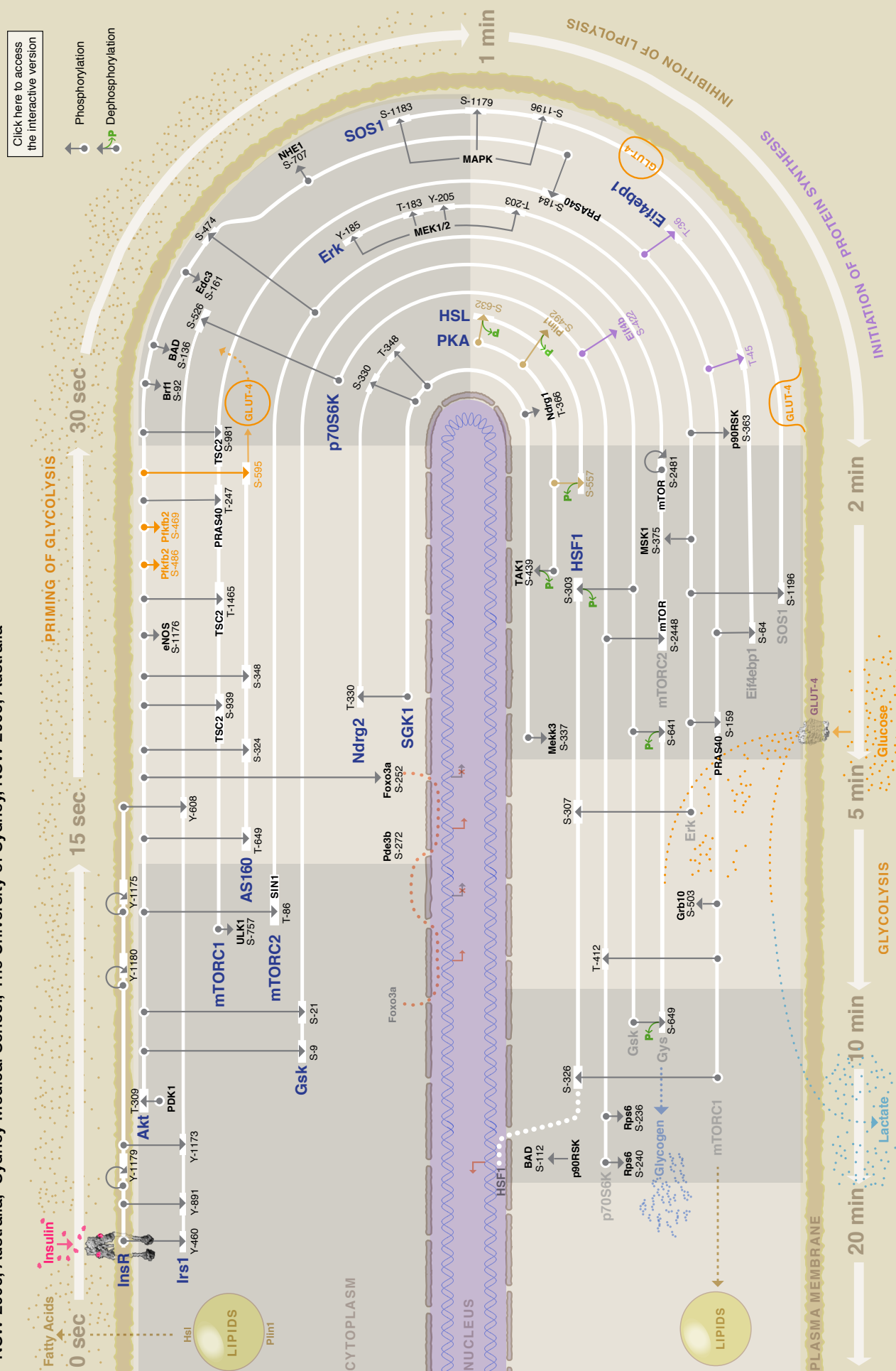
Ilana Kolodkin-Gal, Shugeng Cao, Liraz Chai, Thomas Böttcher, Roberto Kolter, Jon Clardy,* and Richard Losick*

*Correspondence: jon_clardy@hms.harvard.edu (J.C.), losick@mcb.harvard.edu (R.L.)
<http://dx.doi.org/10.1016/j.cell.2015.04.039>

(Cell 149, 684–692; April 27, 2012)

In this article, we reported that norspermidine is produced in aged biofilm cultures of *Bacillus subtilis* and that norspermidine could disassemble and inhibit *B. subtilis* biofilms. Both claims have been challenged by Hobley et al. (2014, Cell 156, 844–854). We have subsequently repeated the experiments and have found that the new results can no longer support our original conclusions. Therefore, the most appropriate course of action is to retract the article. We offer our apologies for these errors and for any inconvenience that they may have caused.

David K.G. Ma,^{1,2} Christian Stolte,³ James R. Krycer,⁴ David E. James,^{4,5} and Seán I. O'Donoghue^{1,3}
¹Garvan Institute of Medical Research, Darlinghurst NSW 2010, Australia, ²School of Computer Science and Engineering, UNSW Sydney,
 NSW 2052, Australia, ³CSIRO, North Ryde NSW 1670, Australia, ⁴Charles Perkins Centre, School of Molecular Bioscience, The University of Sydney,
 NSW 2006, Australia, ⁵Sydney Medical School, The University of Sydney, NSW 2006, Australia



David K.G. Ma,^{1,2} Christian Stolte,³ James R. Krycer,⁴ David E. James,^{4,5} and Seán I. O'Donoghue^{1,3}

¹Garvan Institute of Medical Research, Darlinghurst NSW 2010, Australia

²School of Computer Science and Engineering, UNSW Sydney, NSW 2052, Australia

³CSIRO, North Ryde NSW 1670, Australia

⁴Charles Perkins Centre, School of Molecular Bioscience, The University of Sydney, NSW 2006, Australia

⁵Sydney Medical School, The University of Sydney, NSW 2006, Australia

The insulin/IGF1 signaling pathway (ISP) plays an essential role in long-term health. Some perturbations in this pathway are associated with diseases such as type 2 diabetes (George et al., 2004); other perturbations extend lifespan in worms, flies, and mice (Ziv and Hu, 2011). The ISP regulates many biological processes, including energy storage, apoptosis, transcription, and cellular homeostasis. Such regulation involves precise rewiring of temporal events in protein phosphorylation networks; these events can now be observed in detail using high-throughput mass spectrometry (Humphrey et al., 2013).

To address the challenge of displaying the resulting multi-dimensional data sets, we developed Minardo, a novel strategy for visualizing time-course events in cellular systems (Ma et al., 2013). Beginning with the moment that the adipocyte perceives an increase in extracellular insulin (upper-left), the Minardo layout shows the progressive triggering of nodes in the first 20 min of the ISP cascade, with time flowing in a clockwise direction. Gray arrows indicate phosphorylation or dephosphorylation events, from kinase (arrow stem) to substrate (arrowhead), with position indicating the time of half-maximal change in phosphorylation state. Some proteins (or complexes) play major roles in ISP and hence have events at multiple times; these are represented as white tracks. Residue numbering refers to mouse proteins, as used by Humphrey et al. (2013).

We see that the insulin signal is rapidly transduced via the insulin receptor and the scaffold protein IRS-1 to the serine/threonine kinase Akt, which then targets substrates involved in numerous biological pathways (many involving 14-3-3). One prominent example is glucose disposal: Akt phosphorylates AS160 to promote glucose uptake, via provocation of the translocation of intracellular GLUT4 glucose transporter vesicles to the plasma membrane of the cell (Stöckli et al., 2011), and phosphorylates PFKFB2 to stimulate glycolysis (Deprez et al., 1997). The kinase GSK3 α/β is inhibited by Akt, which in turn derepresses glycogen synthase to facilitate incorporation of the incoming glucose into glycogen. In contrast, Akt phosphorylates PRAS40 and TSC2 to activate mTORC1, a key kinase that drives lipid and amino acid metabolism (Efeyan et al., 2012). Thus, Akt straddles several kinase cascades, evident from the progression of the tracks in the Minardo layout.

Beyond Akt, insulin represses protein kinase A (PKA) by activating Pde3b to hydrolyze cAMP, an allosteric activator of PKA (Kitamura et al., 1999; Onuma et al., 2002). This deactivates downstream substrates such as Hsl and Plin1, leading to reduced lipid breakdown and free fatty-acid release within about 1 min. Thus, insulin signaling acts through several kinases and a multitude of substrates to promote energy storage and prevent energy mobilization (Saltiel and Kahn, 2001).

Cellular metabolism is regulated alongside gene expression, with mTORC1 playing a central role. mTORC1 and its downstream kinase, p70S6K, phosphorylate a range of proteins involved in translation (Rps6, Eif4b, Eif4ebp1) while promoting the heat-shock response protein Hsf1. Insulin signaling also promotes mRNA stability by phosphorylation of Brf1 and Edc3, together stimulating a global increase in protein synthesis. mTORC1 also instructs the cell that nutrients are not limiting by phosphorylating ULK1, which curbs autophagy (Efeyan et al., 2012). Furthermore, insulin signaling prevents apoptosis by inhibiting Bad from binding Bcl and trapping the transcription factor Foxo3a within the cytoplasm. This coordinated regulation ensures that the enormous energetic demands of these various biosynthetic processes are balanced by fuel supply, particularly from glucose metabolism.

This complex network is maintained by crosstalk between key kinases, shown by connections between the corresponding tracks in the Minardo layout. For instance, after Akt is phosphorylated at T308, it phosphorylates SIN1, which activates mTORC2 to phosphorylate Akt at S473, completely activating Akt (Humphrey et al., 2013). In contrast, mTORC1 phosphorylates Grb10, and p70S6K phosphorylates IRS1 and mTOR, together attenuating insulin signaling. Such feedback mechanisms fine-tune the appropriate responses to environmental changes.

In summary, insulin/IGF1 signaling involves several key kinase nodes that target many substrates, thus intertwining metabolism with numerous other biological processes.

For an animated version of this SnapShot, please see <http://www.cell.com/cell/enhanced/odonoghue>.

REFERENCES

- Deprez, J., Vertommen, D., Alessi, D.R., Hue, L., and Rider, M.H. (1997). *J. Biol. Chem.* 272, 17269–17275.
- Efeyan, A., Zoncu, R., and Sabatini, D.M. (2012). *Trends Mol. Med.* 18, 524–533.
- George, S., Rochford, J.J., Wolfrum, C., Gray, S.L., Schinner, S., Wilson, J.C., Soos, M.A., Murgatroyd, P.R., Williams, R.M., Acerini, C.L., et al. (2004). *Science* 304, 1325–1328.
- Humphrey, S.J., Yang, G., Yang, P., Fazakerley, D.J., Stöckli, J., Yang, J.Y., and James, D.E. (2013). *Cell Metab.* 17, 1009–1020.
- Kitamura, T., Kitamura, Y., Kuroda, S., Hino, Y., Ando, M., Kotani, K., Konishi, H., Matsuzaki, H., Kikkawa, U., Ogawa, W., and Kasuga, M. (1999). *Mol. Cell. Biol.* 19, 6286–6296.
- Ma, D.K.G., Stolte, C., Kaur, S., Bain, M., and O'Donoghue, S.I. (2013). *AIP Conf. Proc.* 1559, 185–196.
- Onuma, H., Osawa, H., Yamada, K., Ogura, T., Tanabe, F., Granner, D.K., and Makino, H. (2002). *Diabetes* 51, 3362–3367.
- Saltiel, A.R., and Kahn, C.R. (2001). *Nature* 414, 799–806.
- Stöckli, J., Fazakerley, D.J., and James, D.E. (2011). *J. Cell Sci.* 124, 4147–4159.
- Ziv, E., and Hu, D. (2011). *Ageing Res. Rev.* 10, 201–204.